

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:17:36 : Search time 517.286 Seconds
(without alignments)
54.868 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITVLLILAVLLGSIPLSG.....NKTENLVWGLNLTHQDPKK 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DR seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A_Geneseq.101002.*			
1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1071	100.0	213	21	AA192246		S. pneumoniae S-yn
2	1071	100.0	213	22	AAU37736		Streptococcus pneu
3	1071	100.0	213	22	AAU37943		Streptococcus pneu
4	716.5	66.9	213	23	ABB54294		Lactococcus lactis
5	716	66.9	240	23	ABP26819		Streptococcus poly
6	708	66.1	212	23	ABP26818		Streptococcus poly
7	700	65.4	207	23	ABP30349		Streptococcus poly
8	548	51.2	114	20	AA125561		Bacterial general
9	468	43.7	193	20	AA122580		B. subtilis B-ynes
10	468	43.7	193	21	AA192247		B. subtilis B-ynes

11	463.5	43.3	198	23	ABB48775		Listeria monocytog
12	463.5	43.3	204	23	ABP40394		Staphylococcus epi
13	450.5	42.1	197	22	AAG82881		S. epidermidis ope
14	430.5	40.9	202	22	AAU33809		Staphylococcus aur
15	430.5	40.9	202	22	AAU36544		Staphylococcus aur
16	266.5	24.9	203	22	AAU38164		Salmonella typhi c
17	263.5	24.6	205	22	AAU34710		E. coli cellular p
18	252.5	23.6	205	22	AAU36120		Klebsiella pneumon
19	249.5	23.3	173	21	AAU36120		Klebsiella pneumon
20	244.5	22.8	199	22	AAU35404		Haemophilus influe
21	240	22.4	173	21	AAU35404		Haemophilus influe
22	239.5	22.4	173	21	AAU35404		Haemophilus influe
23	239	22.3	189	22	AAU36206		Neisseria gonorrh
24	213	19.9	220	22	AAU36021		Pseudomonas aerugi
25	211	19.7	262	19	AAU98757		Helicobacter pylor
26	211	19.7	262	19	AAU98757		Helicobacter pylor
27	130	12.1	336	22	ABP26867		Streptococcus poly
28	129	12.0	336	20	AAU14584		Streptococcus poly
29	110.5	10.3	461	23	ABP26853		Streptococcus pneu
30	104.5	9.8	419	23	ABP28128		Streptococcus poly
31	103	9.6	542	22	ABB56844		Drosophila melanog
32	102.5	9.6	641	23	ABP30392		Streptococcus poly
33	102.5	9.6	654	23	ABP28438		Streptococcus poly
34	102	9.5	650	22	AAU37748		Streptococcus pneu
35	101	9.4	342	22	AAG82880		S. epidermidis ope
36	101	9.4	347	23	ABP39115		Staphylococcus epi
37	100.5	9.4	456	22	ABP26867		Amino acid sequenc
38	100	9.3	629	22	ABB40427		Drosophila melanog
39	99	9.2	629	22	AAE00297		Drosophila melanog
40	98.5	9.2	388	23	ABP47596		Listeria monocytog
41	98	9.2	391	23	ABP47596		Listeria monocytog
42	97.5	9.1	523	23	AAE14405		Human neurotransmi
43	97	9.1	179	23	ABP25497		Streptococcus poly
44	97	9.1	332	22	AAH87793		Rat T2R13 amino ac
45	96.5	9.0	377	23	ABP26528		Streptococcus poly

ALIGNMENTS

RESULT 1	
AA192246	
ID	AA192246 standard; Protein; 213 AA.
XX	
AC	AA192246;
XX	
DT	10-AUG-2000 (first entry)
XX	
DE	S. pneumoniae S-ynes polypeptide.
XX	
KW	S-ynes; survival; antibacterial; inhibitor.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	WO200020627-A1.
XX	
PD	13-APR-2000.
XX	
PF	30-SEP-1999; 99WO-US22665.
XX	
PR	30-SEP-1998; 98US-0163445.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Fritz C, Youngman P, Guzman L;
XX	
DR	WPI; 2000-303799/26.
XX	
DR	N-PSDB; AA09181.
XX	
PT	Methods for identifying an antibacterial agent for treating
PT	Streptococcus pneumoniae infections comprises detecting an interaction
PT	between a ynes polypeptide and a test compound
XX	

PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB; AAS55802.

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX

PS Example 3; Seq ID No 13536; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 213 AA:

Query Match 100.0%; Score 1071; DB 22; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.1e-116;
 Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVID 60
 DB 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVID 60
 QY 61 FFKGTATLLPIIFHLQGVSPLIIFGLLAVIGHTFPFIFAGFKGKAVATSGAVIFGPAPIF 120
 DB 61 FFKGTATLLPIIFHLQGVSPLIIFGLLAVIGHTFPFIFAGFKGKAVATSGAVIFGPAPIF 120
 QY 121 CLYLAIFFGALYLSMISLSSVSTASIAAVIGVLLPFLFGFILSNYDSLFIAIILASL 180
 DB 121 CLYLAIFFGALYLSMISLSSVSTASIAAVIGVLLPFLFGFILSNYDSLFIAIILASL 180
 QY 181 IIRHKDNRIKKNKTENLVPMGLNTHQDPK 213
 DB 181 IIRHKDNRIKKNKTENLVPMGLNTHQDPK 213

RESULT 4

ABB54294

ID ABB54294 standard; Protein; 213 AA.

XX ABB54294;

XX 16-MAY-2002 (first entry)

DE Lactococcus lactis protein ykac.

KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis IL1403.
 XX FR2807446-A1.
 XX 12-OCT-2001.
 XX 11-APR-2000; 2000FR-0004630.
 XX 11-APR-2000; 2000FR-0004630.

XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX WPI: 2002-043418/06.
 XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -
 XX Claim 6; SEQ ID No 996; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 213 AA:

Query Match 66.9%; Score 716.5; DB 23; Length 213;
 Best Local Similarity 64.8%; Pred. No. 2.7e-75;
 Matches 138; Conservative 35; Mismatches 39; Indels 1; Gaps 1;
 QY 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVI 59
 DB 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVI 60
 QY 60 DFFKGTATLLPIIFHLQGVSPLIIFGLLAVIGHTFPFIFAGFKGKAVATSGAVIFGPAP 119
 DB 61 DFFKGTATLLPIIFHLQGVSPLIIFGLLAVIGHTFPFIFAGFKGKAVATSGAVIFGPAP 120
 QY 120 FCYLAIFFGALYLSMISLSSVSTASIAAVIGVLLPFLFGFILSNYDSLFIAIILASL 179
 DB 121 FCYLAIFFGALYLSMISLSSVSTASIAAVIGVLLPFLFGFILSNYDSLFIAIILASL 180
 QY 180 LTIIRHKDNRIKKNKTENLVPMGLNTHQDPK 212
 DB 181 LTIIRHKDNRIKKNKTENLVPMGLNTHQDPK 213

RESULT 5

ABP26819

ID ABP26819 standard; Protein; 240 AA.

XX ABP26819;

XX 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 2814.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX Streptococcus pyogenes.
 OS

PN W0200234771-A2.
XX 02-MAY-2002.
XX 29-OCT-2001; 2001WO-GB04789.
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX WPI: 2002-352536/38.
XX N-PSDB; ABN67450.
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX Claim 1; Page 3432; 4525pp; English.
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX SQ Sequence 240 AA;
Query Match 66.9%; Score 716; DB 23; Length 240;
Best Local Similarity 64.8%; Pred. No. 3.6e-75;
Matches 138; Conservative 27; Mismatches 48; Indels 0; Gaps 0;
QY 1 MITIVLILAYLGSIPSGLWIGOVFFQINLREHSGSGTGTNTFRILGKAGMATFVID 60
DB 28 MKLLFTTIAIYLLGSIPTGLWIGQYFVHINLREHSGSGTGTNTFRILGKAGTATLAD 87
QY 61 FFKGTLATLPIIFHLQGVSPFLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGPAPIF 120
DB 88 MFKGTLILPIIFGMSISIAIGTFVAVLGHFTFPFANFKGKAVATSAGVILGAPLY 147
QY 121 CLYLAIFFGALYGLSMISLSSVSTASIAAVIGVLPLFLPGFILSNDSLFIAILALASL 180
DB 148 LFLASIFVLVLYLFSMISLASVSAIVGVLVLTPTPAIFLPLPNVDYFLTFVILLAFI 207
QY 181 IIRHKDNIAIRIKNKTENLVPMGLNLTHODPKK 213
DB 208 IIRHKDNISRIKHHTENLIPWGLNLKQVPPK 240
RESULT 6
ABP26818
ID ABP26818 standard; Protein: 212 AA.
XX AC ABP26818;
XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 2812.
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX W0200234771-A2.
XX 02-MAY-2002.
XX 29-OCT-2001; 2001WO-GB04789.
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX WPI: 2002-352536/38.
XX N-PSDB; ABN67449.
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX Claim 1; Page 3432; 4525pp; English.
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX SQ Sequence 212 AA;
Query Match 66.1%; Score 708; DB 23; Length 212;
Best Local Similarity 63.2%; Pred. No. 2.6e-74;
Matches 134; Conservative 33; Mismatches 45; Indels 0; Gaps 0;
QY 1 MITIVLILAYLGSIPSGLWIGOVFFQINLREHSGSGTGTNTFRILGKAGMATFVID 60
DB 1 MNIIIMIIITAYLGSITQTLWIKYFYQVNLROHSGSGTGTNTFRILGKAGIVTLTID 60
QY 61 FFKGTLATLPIIFHLQGVSPFLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGPAPIF 120
DB 61 ILKGTLATLPIILGITTVSPFFIGFFAIIIGHTFPFIFOAKGKAVATSAGVILGAPSF 120
QY 121 CLYLAIFFGALYGLSMISLSSVSTASIAAVIGVLPLFLPGFILSNDSLFIAILALASL 180
DB 121 FLYLLVFLTLVLYLFSMISLSSITVAVVGLSVLIFPLVGLFILTIDYDWTFTVILMALT 180
QY 181 IIRHKDNIAIRIKNKTENLVPMGLNLTHODPKK 212
DB 181 IIRHKDNIAIRIKRQENLVPPGLNLKQKNK 212

ID AAY22580 standard; Protein; 193 AA.
 AC AAY22580;
 XX
 DT 17-NOV-1999 (first entry)
 XX
 DE B. subtilis B-ynes protein sequence.
 XX
 KW General essential protein; pathogenic bacteria; pathogen; inhibitor;
 KW bacterial growth; B-ynes.
 XX
 OS Bacillus subtilis.
 XX
 PN WO9933871-A2.
 XX
 PD 08-JUL-1999.
 XX
 PF 30-DEC-1998; 98WO-US27918.
 XX
 PR 31-DEC-1997; 97US-0070116.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Youngman P, Fritz C, Murphy C, Guzman L;
 XX
 DR WPI; 1999-430230/36.
 DR N-PSDB; AA220371.
 XX
 PT Streptococcus pneumoniae general essential protein genes and proteins,
 PT useful for identification of antibacterial agents -
 XX
 PS Disclosure; Fig 24; 124pp; English.
 XX
 CC This sequence is the Bacillus subtilis B-ynes protein. B-ynes is
 CC related to the Streptococcus pneumoniae general essential
 CC protein (GEP) gene of the invention. The genes encoding the GEP
 CC polypeptides are useful molecular tools for identifying similar genes in
 CC pathogenic microorganisms, such as pathogenic strains of Bacillus. In
 CC addition, the operons containing genes encoding GEP and the polypeptides
 CC themselves, are useful targets for identifying compounds that are
 CC inhibitors of the pathogens in which the GEP are expressed. Such
 CC inhibitors are useful for inhibiting bacterial growth by being
 CC bacteriostatic or bacteriocidal.
 XX
 SQ Sequence 193 AA;
 Query Match 43.7%; Score 468; DB 20; Length 193;
 Best Local Similarity 48.0%; Pred. No. 2.4e-46;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MTTIVLLIAYLLGSLIPSGLVGQVFFQINLREHSGNGTGTNTFRILGKKGAMATFVID 60
 DB 1 MLTALLIAYLLGSLIPSGLVGKLAGKDIREHSGNLGATNAPFTLVGKAGSVVIAGD 60
 QY 61 FFKGTALPLPIFHQLGVSPLIFGLLAVIGHTFTPIFAGKGGKAVATSAGVIFGAPIF 120
 DB 61 ILKGTALPLPMLHVD-IHPLLAGVFAVLGHVFPFIFAKFGKAVATSGGVLLFYAPLL 119
 QY 121 CLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLPFLFGFILSNYDSLFTATILASL 180
 DB 120 FITWAVVFELTYLTFVSLSSMLTGIYTVI-----YSFFV--HDTYLLIVVLLTTF 170
 QY 181 IIRHKDNIAIRKNKTENLVPW 202
 DB 171 VIYHRANIKRIINKTEPKVKW 192
 RESULT 10
 ID AAY92247
 XX AAY92247 standard; Protein; 193 AA.
 AC AAY92247;
 XX

DT 10-AUG-2000 (first entry)
 XX
 DE B. subtilis B-ynes polypeptide.
 XX
 KW B-ynes; survival; antibacterial; inhibitor.
 XX
 OS Bacillus subtilis.
 XX
 PN WO200020627-A1.
 XX
 PD 13-APR-2000.
 XX
 PF 30-SEP-1999; 99WO-US22665.
 XX
 PR 30-SEP-1998; 98US-0163445.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Fritz C, Youngman P, Guzman L;
 XX
 DR WPI; 2000-303799/26.
 DR N-PSDB; AAA09182.
 XX
 PT Methods for identifying an antibacterial agent for treating
 PT Streptococcus pneumoniae infections comprises detecting an interaction
 PT between a ynes polypeptide and a test compound
 XX
 PS Disclosure; Fig 2; 65pp; English.
 XX
 CC Bacillus subtilis ynes gene is a homologue of Streptococcus pneumoniae
 CC ynes gene, which encodes a polypeptide (S-ynes) essential for survival
 CC for a wide range of bacteria. Identifying an antibacterial agent
 CC comprises contacting a ynes polypeptide (S-ynes) with a test compound and
 CC detecting an interaction of the test compound with the S-ynes polypeptide
 CC which indicates that the compound is an antibacterial agent.
 CC Alternatively, detecting a decrease in function of the polypeptide
 CC contacted with the test compound and determining whether the compound
 CC inhibits growth of bacteria, relative to the growth of bacteria cultured
 CC in the absence of a test compound where inhibition of growth indicates
 CC the compound is an antibacterial agent. Inhibitors of S-ynes function
 CC are useful for treating a Streptococcus pneumoniae infection in mammals.
 XX
 SQ Sequence 193 AA;
 Query Match 43.7%; Score 468; DB 21; Length 193;
 Best Local Similarity 48.0%; Pred. No. 2.4e-46;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MTTIVLLIAYLLGSLIPSGLVGQVFFQINLREHSGNGTGTNTFRILGKKGAMATFVID 60
 DB 1 MLTALLIAYLLGSLIPSGLVGKLAGKDIREHSGNLGATNAPFTLVGKAGSVVIAGD 60
 QY 61 FFKGTALPLPIFHQLGVSPLIFGLLAVIGHTFTPIFAGKGGKAVATSAGVIFGAPIF 120
 DB 61 ILKGTALPLPMLHVD-IHPLLAGVFAVLGHVFPFIFAKFGKAVATSGGVLLFYAPLL 119
 QY 121 CLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLPFLFGFILSNYDSLFTATILASL 180
 DB 120 FITWAVVFELTYLTFVSLSSMLTGIYTVI-----YSFFV--HDTYLLIVVLLTTF 170
 QY 181 IIRHKDNIAIRKNKTENLVPW 202
 DB 171 VIYHRANIKRIINKTEPKVKW 192
 RESULT 11
 ID ABB48775
 XX ABB48775 standard; Protein; 198 AA.
 AC ABB48775;
 XX
 DT 05-FEB-2002 (first entry)
 XX

Matches 89; Conservative 44; Mismatches 60; Indels 17; Gaps 3;

OY 1 MITVILLIAYLLGSLWQVFFQINLREHSGNGTNTFRILGKAGMATFVID 60
 DB 1 MMIIIVMLLSYLLGAFPSGVIGKLFKKDIROFGSGNTGATNSFRVLGRPAGFLVTFELD 60
 OY 61 FFKGTATLLPIIFHLOGVSP-----LIFGLLAVIGHTPPIFAGFKGKAVATSAGV 112
 DB 61 IFKGFITVFFPLWLPVHADGPISFTFTNGLIVGLFAILGHVYPVYLKFOGKAVATSAGV 120
 OY 113 IFGFAPIFCLYLAIIFFGALYLGSMISLSSVTASIAAIVIGVLLFPLFGFILSNYDSLFTA 172
 DB 121 VLGUNPILLILAIFFIVKIFKYSVASIACCVIGSL-----IIQDYILLVVS 173
 OY 173 IILASLIIIRHKDNIARKNTENLVPW 202
 DB 174 FLVSI--ILIIHRNSISIRFGEPEKIKW 201

RESULT 15

AAU36544
 ID AAU36544 standard; Protein: 202 AA.

AC AAU36544;

DT 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #714.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS54403.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 12137; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC

CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 202 AA;

Query Match 40.9%; Score 438.5; DB 22; Length 202;
 Best Local Similarity 42.4%; Pred. No. 7.2e-43;
 Matches 89; Conservative 44; Mismatches 60; Indels 17; Gaps 3;

OY 1 MITVILLIAYLLGSLWQVFFQINLREHSGNGTNTFRILGKAGMATFVID 60

DB 1 MMIIIVMLLSYLLGAFPSGVIGKLFKKDIROFGSGNTGATNSFRVLGRPAGFLVTFELD 60

OY 61 FFKGTATLLPIIFHLOGVSP-----LIFGLLAVIGHTPPIFAGFKGKAVATSAGV 112

DB 61 IFKGFITVFFPLWLPVHADGPISFTFTNGLIVGLFAILGHVYPVYLKFOGKAVATSAGV 120

OY 113 IFGFAPIFCLYLAIIFFGALYLGSMISLSSVTASIAAIVIGVLLFPLFGFILSNYDSLFTA 172

DB 121 VLGUNPILLILAIFFIVKIFKYSVASIACCVIGSL-----IIQDYILLVVS 173

OY 173 IILASLIIIRHKDNIARKNTENLVPW 202

DB 174 FLVSI--ILIIHRNSISIRFGEPEKIKW 201

Search completed: December 26, 2002, 00:49:11

Job time : 513.286 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	548	51.2	114	4	US-09-222-938A-13	Sequence 13, Appl
2	468	43.7	193	4	US-09-222-938A-70	Sequence 70, Appl
3	463.5	43.3	204	4	US-09-134-001C-5239	Sequence 5239, Ap
4	101	9.4	347	4	US-09-134-001C-3960	Sequence 3960, Ap
5	93	8.7	323	4	US-09-134-001C-4635	Sequence 4635, Ap
6	92.5	8.6	405	4	US-09-134-001C-4999	Sequence 4999, Ap
7	91.5	8.5	287	4	US-09-134-001C-5055	Sequence 5055, Ap
8	91.5	8.5	438	2	US-08-677-049-9	Sequence 9, Appli
9	91.5	8.5	518	4	US-09-134-001C-4744	Sequence 4744, Ap
10	89.5	8.4	400	4	US-09-134-001C-2912	Sequence 2912, Ap
11	88	8.2	454	4	US-09-134-001C-3501	Sequence 3501, Ap
12	87.5	8.2	799	4	US-09-165-396-4	Sequence 4, Appli
13	86	8.0	655	4	US-09-245-808-1	Sequence 1, Appli
14	86	8.0	901	4	US-09-134-001C-5389	Sequence 5389, Ap
15	85.5	8.0	190	4	US-09-134-001C-5000	Sequence 5000, Ap
16	85.5	8.0	460	2	US-08-677-049-10	Sequence 10, Appl
17	85.5	8.0	635	2	US-09-014-969-11	Sequence 11, Appl
18	85	7.9	237	2	US-08-849-480A-7	Sequence 7, Appli
19	85	7.9	259	2	US-09-097-759-3	Sequence 3, Appli
20	85	7.9	306	2	US-09-097-759-2	Sequence 2, Appli
21	85	7.9	315	4	US-09-134-001C-5446	Sequence 5446, Ap
22	85	7.9	668	4	US-09-134-001C-3430	Sequence 3430, Ap
23	84.5	7.9	229	4	US-09-227-357-237	Sequence 237, Appl
24	84.5	7.9	530	4	US-09-134-001C-4510	Sequence 4510, Ap
25	84	7.8	372	2	US-08-626-685A-8	Sequence 8, Appli
26	84	7.8	372	4	US-08-993-088A-2	Sequence 2, Appli
27	84	7.8	372	4	US-08-993-424B-2	Sequence 2, Appli

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 00:31:11 : Search time 332.091 Seconds
(without alignments)
61.660 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITIVLLILAYLLGSLIPSG.....NKTENLVPGWLNLTHTQDPKK 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-73:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1071	100.0	213	2 C95098	conserved hypohet
2	1071	100.0	213	2 C97966	conserved hypohet
3	716.5	66.9	213	2 B86747	conserved hypohet
4	468	43.7	193	1 A69892	conserved hypohet
5	463.5	43.3	198	2 AD1235	B. subtilis YneS p
6	459.5	42.9	198	2 AB1598	conserved hypohet
7	441.5	41.2	202	2 A89911	conserved hypohet
8	332	31.0	192	2 E70359	conserved hypohet
9	300.5	28.1	222	1 S75137	hypothetical prote
10	298	27.8	206	2 G83854	hypothetical prote
11	291.5	27.2	198	2 A75324	conserved hypohet
12	282.5	26.4	201	2 AC3593	probable integral
13	282	26.3	202	2 AB2737	conserved hypohet
14	282	26.3	205	2 H97517	hypothetical prote
15	275.5	25.7	198	2 C97208	uncharacterized co
16	275.5	25.7	200	2 C81894	probable integral
17	274	25.6	208	2 B82313	conserved hypohet
18	269.5	25.2	196	2 C72253	conserved hypohet
19	266.5	24.9	203	2 AF0892	probable membrane
20	266	24.8	200	2 H81126	conserved hypohet
21	265.5	24.8	216	2 AB0080	probable membrane
22	263.5	24.6	205	1 A65094	ygiH protein - Esc
23	263.5	24.6	205	2 F91121	hypothetical prote
24	263.5	24.6	202	2 E85966	hypothetical prote
25	256.5	23.9	202	2 D81378	probable integral
26	255.5	23.9	226	2 AC1868	hypothetical prote
27	244.5	22.8	199	2 E64146	hypothetical prote
28	239	22.3	189	2 A83573	conserved hypohet
29	236.5	22.1	257	2 A82894	conserved hypohet

30	236	22.0	218	2 E87552	conserved hypohet
31	227	21.2	224	2 G90548	conserved hypohet
32	223.5	20.9	239	1 C64227	hypothetical prote
33	214.5	20.0	239	2 S73812	hypothetical prote
34	213	19.9	220	2 B71811	hypothetical prote
35	211	19.7	262	1 E64708	conserved hypohet
36	159.5	14.9	203	2 F72273	conserved hypohet
37	156	14.6	559	2 B75477	conserved hypohet
38	144	13.4	195	2 A75295	conserved hypohet
39	122.5	11.4	963	2 AF2119	hypothetical prote
40	113.5	10.6	348	2 T12290	NADH2 dehydrogenas
41	109	10.2	504	2 AB0839	probable membrane
42	107.5	10.0	405	2 B69200	hypothetical prote
43	107	10.0	640	2 D90174	hypothetical prote
44	106	9.9	267	2 AI2794	conserved hypohet
45	106	9.9	310	2 H97573	hypothetical prote

ALIGNMENTS

RESULT 1

C95098
conserved hypothetical protein SP0851 [imported] - Streptococcus pneumoniae (strai
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95098
R:Rettelin, H.; Neilson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtz
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Mo
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoni
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <CUR>
A:Cross-references: GB:AE005672; PIDN:AAK74980.1; PID:g14972323; GSPDB:GN00164; T1
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0851
C:Superfamily: Escherichia coli ygiH protein

Query Match	100.0%	Score 1071;	DB 2;	Length 213;
Best Local Similarity	100.0%;	Pred. No. 1.5e-80;		
Matches 213;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MITIVLLILAYLLGSLIPSGWVFFQINLRHSGSGNTGTTNFRILGKKAGMATFVID 60			
Db 1	MITIVLLILAYLLGSLIPSGWVFFQINLRHSGSGNTGTTNFRILGKKAGMATFVID 60			
QY 61	FFKGTATLPIIFHLOGVSPLIIFGLAVIHTFFIFAGFKGKAVATSAGVIFGPAIF 120			
Db 61	FFKGTATLPIIFHLOGVSPLIIFGLAVIHTFFIFAGFKGKAVATSAGVIFGPAIF 120			
QY 121	CLYLAIIFGALYLSMISLSSVTASIAAVIGVLLPLFGFILSNYDSLFIAIILASL 180			
Db 121	CLYLAIIFGALYLSMISLSSVTASIAAVIGVLLPLFGFILSNYDSLFIAIILASL 180			
QY 181	IIIRHKDNTARIKNKTENLVPGWLNLTHTQDPKK 213			
Db 181	IIIRHKDNTARIKNKTENLVPGWLNLTHTQDPKK 213			

RESULT 2

C97966
conserved hypothetical protein spr0755 [imported] - Streptococcus pneumoniae (str
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, I
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren,

A:Residues: 1-198 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAC99362.1; PID:g16410700; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lml284

C:Superfamily: Escherichia coli ygiH protein

Query Match 43.3%; Score 463.5; DB 2; Length 198;

Best Local Similarity 47.8%; Pred. No. 6.4e-31;

Matches 96; Conservative 38; Mismatches 56; Indels 11; Gaps 4;

QY 4 IVLLILAYLGSLGPSGLWGVFQINLREHSGTGTNTFRILKKGKAGMATFVIDFFK 63

DB 6 ILLSLAYVIGSLGPSGLWGVFQINLREHSGTGTNTFRILKKGKAGMATFVIDFFK 65

QY 64 GTLTLPLPIFHQLQVSP--LIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAPIFC 121

DB 66 GTVATLLPFFQLNVHDFWLLTGAFATIGHSPFAGFRGKAVATSAGVILAYAPLLF 125

QY 122 LYLAIFFGALYLGSMISLSSVTASTAAVIGVLLPFLFGFILSNYDSLFIALLASLI 181

DB 126 VAAALVFLVTLKLSKVSLSMMIGALAAI-LSLF--MG-----DWILLIVACIALFV 176

QY 182 IIRHKDNIARIKNTENLVPW 202

DB 177 IWRHRANITRIRNGEPEPKIKW 197

RESULT 6

AB1598 conserved hypothetical protein, B. subtilis yneS protein homolog lin1323 [imported] - Li

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AB1598

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecher

R: Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthi, H.

D.: Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669.

A:Accession: AB1598

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-198 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC96554.1; PID:g16413796; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin1323

C:Superfamily: Escherichia coli ygiH protein

Query Match 42.9%; Score 459.5; DB 2; Length 198;

Best Local Similarity 46.0%; Pred. No. 1.4e-30;

Matches 93; Conservative 41; Mismatches 55; Indels 13; Gaps 4;

QY 4 IVLLILAYLGSLGPSGLWGVFQINLREHSGTGTNTFRILKKGKAGMATFVIDFFK 63

DB 6 ILLSLAYVIGSLGPSGLWGVFQINLREHSGTGTNTFRILKKGKAGMATFVIDFFK 65

QY 64 GTLTLPLPIFHQLQVSP--LIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAPIFC 120

DB 66 GTVATLLPFFQLNVHDFWLLTGAFATIGHSPFAGFRGKAVATSAGVILAYAPLL 124

QY 121 CLYLAIFFGALYLGSMISLSSVTASTAAVIGVLLPFLFGFILSNYDSLFIALLASLI 180

DB 125 FVAALVFLVTLKLSKVSLSMMIGALAAI-----ISFFMG--DWILLIVACIALF 175

QY 181 IIRHKDNIARIKNTENLVPW 202

DB 176 VIWRHRANITRIRNGEPEPKIKW 197

RESULT 7

AB9911

conserved hypothetical protein Sall187 [imported] - Staphylococcus aureus (strain N

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: AB9911

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu

C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuka, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: AB9758; MUID:21311952; PMID:11418146

A:Accession: AB9911

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-202 <KUR>

A:Cross-references: GB:BA000018; PID:g13701150; PIDN:BA042445.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: Sall187

C:Superfamily: Escherichia coli ygiH protein

Query Match 41.2%; Score 441.5; DB 2; Length 202;

Best Local Similarity 42.9%; Pred. No. 4.1e-29;

Matches 90; Conservative 43; Mismatches 60; Indels 17; Gaps 3;

QY 1 MITIVLLILAYLGSLGPSGLWGVFQINLREHSGTGTNTFRILKKGKAGMATFVID 60

DB 1 MMIIIVMLLSYLTGAPFSGVIGKLFKKDIRQFSGNTGATNSFRVLGRPAGFLVTFD 60

QY 61 FPKGTATLPLPIFHQLQVSP-----LIFGLLAVIGHTFPFAGFKGKAVATSAGV 112

DB 61 IFKGTIVTFVFPMLPVHADGPISTFTNGLIVGLFALGHVVPYILKFOGKAVATSAGV 120

QY 113 IFSPAPIFCYLAIFFGALYLGSMISLSSVTASTAAVIGVLLPFLFGFILSNYDSLFI 172

DB 121 VLGVNPILLILALIFVILKIFKVVYSLASIVAAICCVIGSL-----IIQDYILLVVS 173

QY 173 IILASLIIIRHKDNIARIKNTENLVPW 202

DB 174 FLVSI--ILIIHRSNRIARIFRGEPEPKIKW 201

RESULT 8

E70359

conserved hypothetical protein aq_676 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999

C:Accession: E70359

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70359

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-192 <AQF>

A:Cross-references: GB:AE000701; NID:g2983260; PIDN:AAC06869.1; PID:g2983275; GB:A

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq_676

C:Superfamily: Escherichia coli ygiH protein

Query Match 31.0%; Score 332; DB 2; Length 192;

Best Local Similarity 39.8%; Pred. No. 3.5e-20;

Matches 80; Conservative 39; Mismatches 68; Indels 14; Gaps 6;

QY 1 MITIVLLILAYLGSLGPSGLWGVFQINLREHSGTGTNTFRILKKGKAGMATFVID 60

DB 1 MKALFLVIFAYLLGSITTFGEVIAKL-KGVDLRNVGSGNMGATNVTALGKKGVLVFFLD 59

QY 61 FPKGTATLPLPIFHQLQVSPPLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAP 119

Db 60 FLKGFIPALIAVSKFGIDSWLFTFTGLASVLGHMIPVFFGKGGKGVATAGVVFVAVSPS 119
 QY 120 FCYLAIIFPGALYLGSMISLSSVTASIAAIVGVLLPFLFGFILSNY--DSLFAIAILAL 177
 Db 120 VALFSLVWLGIFGLKRYVSLASITATISAF-----LFLV-AGYPVNVLFMAIV--I 169
 QY 178 ASLIIIRHKDNIARIKNKTEN 198
 Db 170 GLLIYHRENNIRLLTGREH 190
 RESULT 9
 S75137
 hypothetical protein sll1973 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: S75137
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S75137
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-222 <KAN>
 A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7999.1; PID:g165308
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: *Escherichia coli* ygiH protein
 Query Match 28.1%; Score 300.5; DB 1; Length 222;
 Best Local Similarity 34.2%; Pred. No. 1.5e-17;
 Matches 75; Conservative 37; Mismatches 66; Indels 41; Gaps 6;
 QY 1 MITIVLLIAYLIGSPGLWIGQVFFQINLREHSGNTGTTTFRILGKKGKAGMATFVID 60
 Db 6 LILCLLLIYLGSIPTGYLAGKLLIGDIREHSGKSTGATNFTLGRKPAIAVLAD 65
 QY 61 FFKGTATLPIIFHLOGVSPLI-----FGLLAVIGHTFPFAGKGGKAVATS 109
 Db 66 ISKGVMAVALVRAIYSDWLPALPAWONWLTLCGVAIVLVGHKSIFLKFSGKSVATS 125
 QY 110 AGVIFGAPIFCLYLAI-----IFGALYLGSMISLSSVTASIAAIVGVLL-----FPL 158
 Db 126 LGVLF-----MLNIWLALGTALTELTFTVIFTRIVSLSSIVAAI-AVNGIALALQPPYLA 180
 QY 159 FGILSNYDSLFTAILALASLIIIRHKDNIARIKNKTE 197
 Db 181 FTFLAGMY-----VIVRHTNIERILQGE 205
 RESULT 10
 G83854
 hypothetical protein BH1639 [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: G83854
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: G83854
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-206 <STO>
 A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA805358.1; GSPDB:GN0
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH1639
 C:Superfamily: *Escherichia coli* ygiH protein

Query Match 27.8%; Score 298; DB 2; Length 206;
 Best Local Similarity 33.0%; Pred. No. 2.2e-17;
 Matches 65; Conservative 49; Mismatches 75; Indels 8; Gaps 4;
 QY 5 VLLIALLIGSIPSGLWIGQVFFQINLREHSGNTGTTTFRILGKKGKAGMATFVIDFFKG 64
 Db 6 LIIWIGSYLGSVSFSYIIAKIKKVDIROHSGNAGATNTLRVLGPGPVTAVLLDLIKG 65
 QY 65 TIATLPIIFHLOGVS--PLIFGLLAVIGHTFPFAGKGGKAVATSAGVIFGAPIFCL 122
 Db 66 VIAVVTVOLTPODGFHFAAAAGIAAIGHNWPIYVGFRCGKVATTIGVLASLVLAAV 125
 QY 123 YLAIIFGALYLGSMISLSSVT-ASTAAVIGVLLFLPFGFILSNYDSLFTAILALASLI 181
 Db 126 LAGVIAIGSIVWTRYVSLGSLLEVTALLAVLSQWFGYPVA---YIVLTIIVAILSM- 181
 QY 182 IIRHKDNIARIKNKTEN 198
 Db 182 -WHRNSIORLLSGTEN 197
 RESULT 11
 A75324
 conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
 C:Species: *Deinococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: A75324
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson,
 M.; Shen, M.; Vanathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski,
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: A75324
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-198 <WHI>
 A:Cross-references: GB:AE002039; GB:AE000513; NID:g6459810; PIDN:AAF11571.1; PID:g
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2021
 A:Map position: 1
 C:Superfamily: *Escherichia coli* ygiH protein
 Query Match 27.2%; Score 291.5; DB 2; Length 198;
 Best Local Similarity 34.5%; Pred. No. 7.4e-17;
 Matches 68; Conservative 39; Mismatches 81; Indels 9; Gaps 3;
 QY 1 MITIVLLIAYLIGSPGLWIGQVFFQINLREHSGNTGTTTFRILGKKGKAGMATFVID 60
 Db 3 LTALLALLSYLIGAIAPAAALARA-RGVDIRKVGSGNSGATNVLRLSLGKGPALLVASED 61
 QY 61 FFKGTATLPIIFHLOGVSPLIFGLLAVIGHTFPFAGKGGKAVATSAGVIFGAPIF 120
 Db 62 ILKGVLAVALRALGSAEWAALCGVLAVIGHNFSPFLAFRCGKGVATSGFVIAIDLPVL 121
 QY 121 CLYLAIIFGALYLGSMISLSSVTASIAAIVGVLLPFLFGFILSNYDSLFTAILALASL 180
 Db 122 GLTFTVLATACWMLTRFVSAGSIMGAFIAGALVVLVLP-----RPTWDR---AAVLFLAAL 173
 QY 181 IIRHKDNIARIKNKTE 197
 Db 174 LVWQHRENIKRLQAGTE 190
 RESULT 12
 AC3593
 probable integral membrane protein [imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AC3593
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mufjer, C.; Los, T.; Iva

.. Mazur, M.; Coltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitens*
A:Reference number: AB3252; PMID:11756688
A:Accession: AC3593
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-201 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53910.1; PID:gl7984852; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BWE110668
A:Map position: 11
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 26.4%; Score 282.5; DB 2; Length 201;
Best Local Similarity 36.6%; Pred. No. 4.5e-16;
Matches 71; Conservative 31; Mismatches 81; Indels 11; Gaps 3;

QY 7 LILAYLLGSPGLWIGOVFFQINLRHSGNVTGTTNFRILGKKAGMATFVIDFFKGTLL 66
DB 15 LIFGYVLGSPGLILTRAGLDVRAIGSGNIGATNLTGNKKLAATLLDLALGTA 74
QY 67 ATLLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFGKGVATSGVIFGFAPIFCLYLAI 126
DB 75 AAL--TAHFQONAAAGFAGFIGHLFPWVGKGVATYGLVGLAWAGALVFAA 132
QY 127 IFFGALYLGSMISLSSVTASIAAVIGVLLFPFLFGFILSNVDSLFIAIILALASLIIRHK 186
DB 133 AWIVTALLARYSSLSALVASL--VPIALY-----SRGNQALAAFAIMTVIVETIKR 183
QY 187 DNIARIKNKTENLV 200
DB 184 ANISRLNGTESKI 197

RESULT 13
AB2737
conserved hypothetical protein Atul306 [imported] - *Agrobacterium tumefaciens* (strain C58)
C:Species: *Agrobacterium tumefaciens*
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB2737
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42312.1; PID:gl7739714; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul306
A:Map position: circular chromosome
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 26.3%; Score 282; DB 2; Length 202;
Best Local Similarity 35.1%; Pred. No. 4.5e-16;
Matches 68; Conservative 40; Mismatches 74; Indels 12; Gaps 4;

QY 8 ILAYLLGSPGLWIGOVFFQINLRHSGNVTGTTNFRILGKKAGMATFVIDFFKGTLLA 67
DB 16 LIGYLLGSPGLILTRMAGLDGVRKIGSGNIGATNLTGNKKLAATLLDLALGTA 75
QY 68 TLLPIIFHLOGV-SPLIFGLLAVIGHTFTPIFAGFGKGVATSGVIFGFAPIFCLYLAI 126
DB 76 VL--VANALMGYEASLVAGFFAFLGHLFPVWLFGKGVAVYIGVLLGAAPLMLAFAL 133

QY 127 IFFGALYLGSMISLSSVTASIAAVIGVLLFPFLFGFILSNVDSLFIAIILALASLIIRHK 186
DB 134 IWLATAFIRYSSLS-----ALLAMLIIPVALVGLPEKTA MLVTLLSVISM--WKHR 184
QY 187 DNIARIKNKTENLV 200
DB 185 ENIRRLMAGTESRI 198

RESULT 14
H97517
hypothetical protein AGR_C2402 [imported] - *Agrobacterium tumefaciens* (strain C58)
C:Species: *Agrobacterium tumefaciens*
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97517
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Go
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markel
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacteri*
A:Reference number: A97359; PMID:11743194
A:Accession: H97517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87097.1; PID:gl5156359; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C2402
A:Map position: circular chromosome
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 26.3%; Score 282; DB 2; Length 205;
Best Local Similarity 35.1%; Pred. No. 4.5e-16;
Matches 58; Conservative 40; Mismatches 74; Indels 12; Gaps 4;

QY 8 ILAYLLGSPGLWIGOVFFQINLRHSGNVTGTTNFRILGKKAGMATFVIDFFKGTLLA 67
DB 19 LIGYLLGSPGLILTRMAGLDGVRKIGSGNIGATNLTGNKKLAATLLDLALGTA 78
QY 68 TLLPIIFHLOGV-SPLIFGLLAVIGHTFTPIFAGFGKGVATSGVIFGFAPIFCLYLAI 126
DB 79 VL--VANALMGYEASLVAGFFAFLGHLFPVWLFGKGVAVYIGVLLGAAPLMLAFAL 136
QY 127 IFFGALYLGSMISLSSVTASIAAVIGVLLFPFLFGFILSNVDSLFIAIILALASLIIRHK 186
DB 137 IWLATAFIRYSSLS-----ALLAMLIIPVALVGLPEKTA MLVTLLSVISM--WKHR 187
QY 187 DNIARIKNKTENLV 200
DB 188 ENIRRLMAGTESRI 201

RESULT 15
C97208
uncharacterized conserved membrane protein, YgiH/UPF0078 family CAC2500 [imported]
C:Species: *Clostridium acetobutylicum*
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: C97208
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteri
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: C97208
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80454.1; PID:gl5025522; GSPDB:GN00168
A:Experimental source: *Clostridium acetobutylicum* ATCC824
C:Genetics:
A:Gene: CAC2500
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 25.7%; Score 275.5; DB 2; Length 198;

Best Local Similarity 32.0%; Pred. No. 1.5e-15;
Matches 66; Conservative 39; Mismatches 82; Indels 19; Gaps 4;

QY	1	MITIVILLIAYLLGSLPGLWICQVFFQINLREHSGNTGTTNTFRILCKKAGMATFVID	60
Db	2	IIVITVWVSFLGSGIPTGLITKLSGIDVRTKSGNIGSTNVKRVAGTKISMITQVMD	61
QY	61	FFKGTLATLLPIIFHLQGVSP-----IFGLLAVIGHTPEIFAGFKGKAVATSAGVI	113
Db	62	ILKGIIPVLLCMLIASKIKLPISTSMYLSIIVIAVILGHDTPTFLGNGKGVTTVGAP	121
QY	114	FGPAIFCLYLAIFFGALYLGSMISLSSVTASIAAVIG--VLLFPLFGFILSNYDSLFI	171
Db	122	FLAPAAVLAVAGAVYFVLRFTKIVSIKSIAGVITMPACIALRLPI-----EITV	172
QY	172	AIILALASLIIRHKDNIARIKNTKTE	197
Db	173	CAIIA-CGLLIRHKDNLRLVNEE	197

Search completed: December 26, 2002, 01:27:05
Job time : 333.091 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 01:16:25 ; Search time 89.1872 Seconds
(without alignments)
41.350 Million cell updates/sec

Title: US-10-068-080-1

Perfect score: 1071

Sequence: 1 MITVILLIAYLGISPSGL.....NKTENLVPLGNLTHQDPKK 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 109717 seqs, 17314136 residues

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2.6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2.6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2.6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2.6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2.6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2.6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2.6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2.6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2.6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2.6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2.6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2.6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2.6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2.6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1071	100.0	213	10	US-09-815-242-13329 Sequence 13329, A
2	1071	100.0	213	10	US-09-815-242-13536 Sequence 13536, A
3	1071	100.0	213	12	US-10-068-080-1 Sequence 1, Appli
4	468	43.7	193	12	US-10-068-080-3 Sequence 3, Appli
5	443.5	41.4	202	10	US-09-823-246-2 Sequence 2, Appli
6	438.5	40.9	202	10	US-09-815-242-5305 Sequence 5305, Ap
7	438.5	40.9	202	10	US-09-815-242-12137 Sequence 12137, A
8	266.5	24.9	203	10	US-09-815-242-13757 Sequence 13757, A
9	263.5	24.6	205	10	US-09-815-242-10303 Sequence 10303, A
10	252.5	23.6	205	10	US-09-815-242-11713 Sequence 11713, A
11	244.5	22.8	199	10	US-09-815-242-10997 Sequence 10997, A
12	239	22.3	189	10	US-09-815-242-11799 Sequence 11799, A
13	213	19.9	220	10	US-09-815-242-11614 Sequence 11614, A
14	211	19.7	252	10	US-09-815-242-11453 Sequence 11453, A
15	105.5	9.9	434	10	US-09-778-927A-79 Sequence 79, Appli
16	102	9.5	650	10	US-09-815-242-13341 Sequence 13341, A
17	100.5	9.4	456	10	US-09-795-693-2 Sequence 2, Appli
18	97.5	9.1	466	10	US-09-778-927A-78 Sequence 78, Appli
19	97.5	9.1	466	10	US-09-778-927A-81 Sequence 81, Appli

20	95.5	8.9	337	10	US-09-815-242-11246 Sequence 11246, A
21	95	8.9	769	10	US-09-815-242-10575 Sequence 10575, A
22	95	8.9	782	10	US-09-815-242-4893 Sequence 4893, Ap
23	94.5	8.8	382	10	US-09-795-693-23 Sequence 23, Appli
24	93	8.7	452	10	US-09-815-242-13983 Sequence 13983, A
25	90.5	8.5	220	9	US-09-860-670-129 Sequence 129, App
26	90.5	8.5	220	9	US-09-764-868-949 Sequence 949, App
27	90.5	8.5	220	9	US-09-764-868-1231 Sequence 1231, Ap
28	88.5	8.3	428	10	US-09-741-869-394 Sequence 394, App
29	88.5	8.3	1594	10	US-09-971-121-4 Sequence 4, Appli
30	88.5	8.3	1642	10	US-09-971-121-2 Sequence 2, Appli
31	88.5	8.3	1642	12	US-10-090-458-5 Sequence 5, Appli
32	88	8.2	760	10	US-09-833-017-26 Sequence 26, Appli
33	87	8.1	264	10	US-09-815-242-13439 Sequence 13439, A
34	87	8.1	264	10	US-09-815-242-13439 Sequence 13439, A
35	87	8.1	1624	12	US-10-090-454-2 Sequence 2, Appli
36	86.5	8.1	246	10	US-09-815-242-11377 Sequence 11377, A
37	86.5	8.1	418	10	US-09-815-242-12003 Sequence 12003, A
38	86.5	8.1	453	10	US-09-815-242-10872 Sequence 10872, A
39	85.5	8.0	448	10	US-09-815-242-12421 Sequence 12421, A
40	85.5	8.0	535	10	US-09-795-693-20 Sequence 20, Appli
41	85.5	8.0	584	10	US-09-815-242-5716 Sequence 5716, Ap
42	85.5	8.0	584	10	US-09-815-242-12476 Sequence 12476, A
43	85.5	8.0	697	10	US-09-881-752A-76 Sequence 76, Appli
44	85.5	8.0	1638	12	US-10-090-458-2 Sequence 2, Appli
45	84	7.8	391	10	US-09-815-242-11566 Sequence 11566, A

ALIGNMENTS

RESULT 1
US-09-815-242-13329
; Sequence 13329, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13329
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13329
Query Match 100.0%; Score 1071; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.8e-95;

Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITIVLLIAYLGLSGIPSGLVGQVFFQINLREHSGNLTGTTNTRILGKKGAGMATFVID 60
DB 1 MITIVLLIAYLGLSGIPSGLVGQVFFQINLREHSGNLTGTTNTRILGKKGAGMATFVID 60

QY 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIFGAPIF 120
DB 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIFGAPIF 120

QY 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180
DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180

QY 181 IIRHKDNIAIKNKNTENLVPWGLNTHODPKK 213
DB 181 IIRHKDNIAIKNKNTENLVPWGLNTHODPKK 213

RESULT 2
US-09-815-242-13536
; Sequence 13536, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13536
; LENGTH: 213
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13536

Query Match 100.0%; Score 1071; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.8e-95;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITIVLLIAYLGLSGIPSGLVGQVFFQINLREHSGNLTGTTNTRILGKKGAGMATFVID 60
DB 1 MITIVLLIAYLGLSGIPSGLVGQVFFQINLREHSGNLTGTTNTRILGKKGAGMATFVID 60

QY 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIFGAPIF 120
DB 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIFGAPIF 120

QY 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180
DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180

DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180

QY 181 IIRHKDNIAIKNKNTENLVPWGLNTHODPKK 213
DB 181 IIRHKDNIAIKNKNTENLVPWGLNTHODPKK 213

RESULT 3
US-10-068-080-1
; Sequence 1, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 213
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-1

Query Match 100.0%; Score 1071; DB 12; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.8e-95;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITIVLLIAYLGLSGIPSGLVGQVFFQINLREHSGNLTGTTNTRILGKKGAGMATFVID 60
DB 1 MITIVLLIAYLGLSGIPSGLVGQVFFQINLREHSGNLTGTTNTRILGKKGAGMATFVID 60

QY 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIFGAPIF 120
DB 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIFGAPIF 120

QY 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180
DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLFGLFGLSNYDSLFIAIILALASL 180

QY 181 IIRHKDNIAIKNKNTENLVPWGLNTHODPKK 213
DB 181 IIRHKDNIAIKNKNTENLVPWGLNTHODPKK 213

RESULT 4
US-10-068-080-3
; Sequence 3, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 193
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-3

Query Match 43.7%; Score 468; DB 12; Length 193;
Best Local Similarity 48.0%; Pred. No. 8.7e-38;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

Oy 1 MITIVLLIAYLIGSLIPSGWVFFOINLREHSGNGTGTNTFRILGKAGMATFVID 60
Db 1 MIUALLIAYLIGSLIPSGWVFFOINLREHSGNGTGTNTFRILGKAGSVVIAGD 60
Oy 61 FPKGTALTLPIIFHLOGVSPFIIFAGLAVIGTFFIFAGKGGKAVATSAGVIFGAPIF 120
Db 61 ILKGTALATPLMHVD-IHPLLACGFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLL 119
Oy 121 CLYLAIIFGALYIGSMISLSVTSIAAIVGVLLFPLFGFILTSDYSLFTAIILALASL 180
Db 120 FITWAVFFIYLFYKFSLSMLGIYTVI-----YSFV--HDTYLLIVVTLTIF 170
Oy 181 IIRHKDNIARIKNTENLVW 202
Db 171 VIYHRANIKRIINKTEPKVKW 192

RESULT 5
US-09-823-246-2
; Sequence 2, Application US/09823246
; Patent No. US20020058789A1
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Warren, Patrick V.
; APPLICANT: Slyvester, Daniel R.
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: ynes
; FILE REFERENCE: GM2001
; CURRENT APPLICATION NUMBER: US/09/823,246
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,496
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 202
; TYPE: PRG
; ORGANISM: Staphylococcus aureus
US-09-823-246-2

Query Match 41.4%; Score 443.5; DB 10; Length 202;
Best Local Similarity 42.9%; Pred. No. 2e-35;
Matches 90; Conservative 43; Mismatches 60; Indels 17; Gaps 3;

Oy 1 MITIVLLIAYLIGSLIPSGWVFFOINLREHSGNGTGTNTFRILGKAGMATFVID 60
Db 1 MMIIVLLLSYLIGAFPSGVIGLFFKKDIQFGSGNTGATNSFVLGRPAGFLVTFD 60
Oy 61 FPKGTALTLPIIFHLOGVSP-----LIFGLAVIGTFFIFAGKGGKAVATSAGV 112
Db 61 IFKGTITVFFPLWLVHADGPSTFTFTNGLVGLFVAILGHVVPVYLKFGGKAVATSAGV 120
Oy 113 IFGAPICFLYLAIFFGALYIGSMISLSVTSIAAIVGVLLFPLFGFILTSDYSLFIA 172
Db 121 VLGVNPIILLIATIFFIYKIVYSLASIAAICCVIGSL-----IIOYVILLVVS 173
Oy 173 IILASLIIRHKDNIARIKNTENLVW 202
Db 174 FLVSI--ILIRHRSNIIRFGEPEKIKW 201

RESULT 6
US-09-815-242-5305
; Sequence 5305, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5305
; LENGTH: 202
; TYPE: PRG
; ORGANISM: Staphylococcus aureus
US-09-815-242-5305

Query Match 40.9%; Score 438.5; DB 10; Length 202;
Best Local Similarity 42.4%; Pred. No. 6e-35;
Matches 89; Conservative 44; Mismatches 60; Indels 17; Gaps 3;
Oy 1 MITIVLLIAYLIGSLIPSGWVFFOINLREHSGNGTGTNTFRILGKAGMATFVID 60
Db 1 MMIIVLLLSYLIGAFPSGVIGLFFKKDIQFGSGNTGATNSFVLGRPAGFLVTFD 60
Oy 61 FPKGTALTLPIIFHLOGVSP-----LIFGLAVIGTFFIFAGKGGKAVATSAGV 112
Db 61 IFKGTITVFFPLWLVHADGPSTFTFTNGLVGLFVAILGHVVPVYLKFGGKAVATSAGV 120
Oy 113 IFGAPICFLYLAIFFGALYIGSMISLSVTSIAAIVGVLLFPLFGFILTSDYSLFIA 172
Db 121 VLGVNPIILLIATIFFIYKIVYSLASIAAICCVIGSL-----IIOYVILLVVS 173
Oy 173 IILASLIIRHKDNIARIKNTENLVW 202
Db 174 FLVSI--ILIRHRSNIIRFGEPEKIKW 201

RESULT 7
US-09-815-242-12137
; Sequence 12137, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12137
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12137

Query Match      40.9%; Score 438.5; DB 10; Length 202;
Best Local Similarity 42.4%; Pred. No. 6e-35;
Matches 89; Conservative 44; Mismatches 60; Indels 17; Gaps 3;

Qy 1 MTTIVLLIAYLLGSLPGWIGVFFQINLREHSGNGTGTNTFRILGKKGKAGMATFVID 60
   1:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MMTIVLLSLILGAPSGFVIGKLFKRDIFGSGNGTATNSRVLGRPAGFLVTFLD 60
   1:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

Qy 61 FFKGTATLLPIIFHLOGVSP-----LIFGLLAVICHTFFPIFAGFGKGVATSGV 112
   11:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 61 IPKGFITVFFPLWLVHADGPSTFTFTNGLVLGLFALIGHVYVPVLYKFGGKAVATSGV 120
   1:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

Qy 113 IFGPAPICLYLAITFFGALYLGSMISLSVSTASTAAVIGVLLPFLFGILSNYDSLPIA 172
   1:||||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 121 VLGVNPILLILAIIFVILKIFKXVLSASIAAICCVIGSL-----IIQDYILLVVS 173
   1:||||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

Qy 173 IILASLIIIRHKDNIARIKNTENLVPM 202
   1:||||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 174 FLVSI--ILIIHRNSIRIFRGEPEPKIW 201
   1:||||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

RESULT 8
US-09-815-242-13757
; Sequence 13757, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10303
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10303

Query Match      24.6%; Score 263.5; DB 10; Length 205;
Best Local Similarity 36.6%; Pred. No. 3.1e-18;
Matches 74; Conservative 36; Mismatches 71; Indels 21; Gaps 9;

Qy 6 LLIILAYLLGSLPGWIGVFFQINLREHSGNGTGTNTFRILGKKGKAGMATFVIDFKGT 65
   1:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 8 MILFAYLCGSISSAILVCRITAGLPDPRESSGNGCATNLRIIGKGAVALIFDILKG- 66
   1:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

Qy 66 LATLLPII-FHLOGVSPLIFGLLAV---IGHTFFIFAGFGKGVATSGVIFGFAPIFC 121
   1:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 67 ---MLPVMGAYALCVTPFWLGLIAIAACLGHIHVPVFFGKGGKVATAGAI---API-G 119
   1:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

Qy 122 LYLAITFFGALYLGSMISLSVSTASTAAVIGVLLPFLFGILSNYDSLFTAILALASLI 181
   1:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 120 WDLTGVNAGTWLLTVLLSGYS---SLGAIVSALIAP---FYVMWFKPQFTFPVMSLSCLI 173
   1:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

Qy 182 IIRHKDNIARIKNTENLVPM 202
   1:||||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 174 LRFHHDNIQRLWRROETKI-W 193
   1:||||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

RESULT 9
US-09-815-242-10303
; Sequence 10303, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10303
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10303

Query Match      24.6%; Score 263.5; DB 10; Length 205;
Best Local Similarity 36.6%; Pred. No. 3.1e-18;
Matches 74; Conservative 36; Mismatches 71; Indels 21; Gaps 9;

Qy 6 LLIILAYLLGSLPGWIGVFFQINLREHSGNGTGTNTFRILGKKGKAGMATFVIDFKGT 65
```

Db 8 MLTAYLCGSISSAILVCRGLGDPRTSGSGNPGATNVLRIKCGKAAGAVLFDVLKG- 66
Qy 66 LATLLPI--IFHLOGVSLIFGLLAV---IGHTFPPIAGKGGKAVATSGVIFGPAIF 120
Db 67 ---MLPVNGAYEL-GVSPFWLGLTIAACLGHIWPFVFGKGGKGVATAGAI---API- 118
Qy 121 CLYLAIFFGALYLGSMISLSSVASTAAVIGVLLFPLFGFILSNYDSLFATIALASL 180
Db 119 GWDLTGVWAGTWLTVLLSGYS---SLGAIVSALIAP---FYVWFKPQPTFPVMSLSCL 172
Qy 181 IIRHKDNIARIKKNKTENLVPW 202
Db 173 ILLRHDNIORLWRROETKI-W 193

RESULT 10

US-09-815-242-11713
; Sequence 11713, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA 011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11713

; LENGTH: 205

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-815-242-11713

Query Match 23.6%; Score 252.5; DB 10; Length 205;

Best Local Similarity 35.6%; Pred. No. 3.4e-17;

Matches 72; Conservative 40; Mismatches 69; Indels 21; Gaps 9;

Qy 6 LLILAYLIGSPISGLTWIGOVFFQINLREHSGSNTGTTNTRFILCKKAGMATFFIDFKGT 65
Db 8 LVLYLACGSISSAILVCRGLGDPRTSGSGNPGATNVLRIKCGKAAGAVLFDVLKG- 66
Qy 66 LATLLPI--IFHLOGVSLIFGLLAV---IGHTFPPIAGKGGKAVATSGVIFGPAIF 120
Db 67 ---MLPVNGAWAL-GLTFPFWLGLVAIAACVGIWPFVFFHFRGGKGVATAGAI---API- 118
Qy 121 CLYLAIFFGALYLGSMISLSSVASTAAVIGVLLFPLFGFILSNYDSLFATIALASL 180
Db 119 GWDLTGVWAGTWLTVLLSGYS---SLGAIVSALIAP---FYVWFKPQPTFPVMSLSCL 172
Qy 181 IIRHKDNIARIKKNKTENLVPW 202

Db 173 ILLRHDNIORLWRROESKI-W 193

RESULT 11

US-09-815-242-10997

; Sequence 10997, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA 011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10997

; LENGTH: 199

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-815-242-10997

Query Match 22.8%; Score 244.5; DB 10; Length 199;

Best Local Similarity 33.0%; Pred. No. 1.9e-16;

Matches 68; Conservative 41; Mismatches 78; Indels 19; Gaps 7;

Qy 1 MITIVLLIAYLLGSPISGLTWIGOVFFQINLREHSGSNTGTTNTRFILCKKAGMATFFID 60
Db 3 LFALFYMLFAYLLGSISSAILICRIAGLPDPGRONGSHNPGATNVLRIKGRKSALAVLIFD 62

Qy 61 FFKGTATLLPI-IFHLOGVSLIFGLLAV---IGHTFPPIAGKGGKAVATSGVIFGF 116

Db 63 MLKG---MIPVWAGYLLGTQFELGNVALGACLGHIFFIPIFFQFKGKGVATAGAI--- 115

Qy 117 APIFCYLAIIFFGALYLGSMISLSSVASTAAVIGVLLFPLFGFILSNYDSLFATIALA 176

Db 116 APISWAVAGSMFGTWIF---VELVSGYSSLSAVISALLVPFYVWVWFKPEFTFPVALV-- 169

Qy 177 LASLIIRHKDNIARIKKNKTENLVPW 202

Db 170 -CCLLIYRHDNIORLWRGOEDKV-W 193

RESULT 12

US-09-815-242-11799

; Sequence 11799, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11799
LENGTH: 189
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11799

Query Match 22.3%; Score 239; DB 10; Length 189;
Best Local Similarity 34.2%; Pred. No. 6.1e-16;
Matches 65; Conservative 28; Mismatches 77; Indels 20; Gaps 3;
QY 5 VLLILAYLGSIPSGMLWGVFQINLRHSGSGNTGTTNTFRILGKAGKAGATVDFPKG 64
DB 4 LLAILAYLGSLSFVALLSFWGFTQDPRASGSGNPGCATNMLRVAGKKLAILTLGSDVGK 63
QY 65 TLATLPIIFHLQGVSLIFGLAVIGHTPPIFAGKGGKAVATSAGVIFGFAPICLYL 124
DB 64 LLPVLVARGLGVMEANWYGIATVGHLYPLFNFPGGKGVATAGMLGLYPPAVLLA 123
QY 125 AIIFFGALYLSMISLSSVTSIAAVIGVLLFPLF-----GFILSNYDSLFIAIILAS 179
DB 124 AAALLTFKLSRTSSLASLAVT-----PLTLPLAAQOPGALLP-----MTVLTG 168
QY 180 LIIRHKDNI 189
DB 169 LIVVHRANL 178

RESULT 13
US-09-815-242-11614
Sequence 11614, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16

PRIOR APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11614
LENGTH: 220
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11614
Query Match 19.9%; Score 213; DB 10; Length 220;
Best Local Similarity 26.6%; Pred. No. 2.2e-13;
Matches 58; Conservative 45; Mismatches 87; Indels 28; Gaps 6;
QY 2 ITIVLLILAYLGSIPSGMLWGVFQINLRHSGSGNTGTTNTFRIL-----GKKAGH 54
DB 11 INVIFTLGLYIGGIPGYALMKIFGYMDITKIGSGGIGATNVLRLQSKGVSNKQMAL 70
QY 55 ATFFVIDFFKGTATLPIIFHLQGVSLIFGLAVIGHTPPIFAGKGGKAVATSAGVIF 114
DB 71 LVLLDLFKGMFAVFLSKLFGLDYSLQNWVAIASILGHCHYSPFLNFGKGVSTINGSVV 130
QY 115 GFAPICLYLAIIFFGALYLSMISLSSVTSIAAVIG-----VLLF-----PLFGFI 162
DB 131 LLPIESLIGLTVWF---FVGKVLKIS---SLASILGVTATVLIFFVPMHHPDSVNI 183
QY 163 LSNWDSLFTAILALASLIIIRHKONIAIKNTENLV 200
DB 184 LREVGOTPMVLIFITL--IKHAGNIFNLTKGKRV 219

RESULT 14
US-09-815-242-11453
Sequence 11453, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11453
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11453

Query Match 19.7%; Score 211; DB 10; Length 262;
Best Local Similarity 27.1%; Pred. No. 4.2e-13;
Matches 56; Conservative 44; Mismatches 79; Indels 28; Gaps 6;
QY 2 ITIVLILAYLIGSIPSGLWIGQVFOINLREHSGNGTGTNTFRIL-----GKKAGM 54
DB 53 INVITFLGGLGGPFGYALMKIFGMDITKIGSGGIGATNLRALOSKGVSNKQMAL 112
QY 55 ATFVIDFFKGTATLPIIFHLOGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIF 114
DB 113 LVLILDPRKMEAVFLSKFLGLDYSLOWMVAIASILGHCSPLNFGNGKGVSTIMGSV 172
QY 115 GFAPICFLYLAIFFGALYLGSMISLSSVTASIAAIG-----VLLF-----PLFGFI 162
DB 173 LLIPIESLGLTWTF---FVGKVLKIS-----SLASILGVGTATVLIFFPYMHIPDSVNI 225
QY 163 LSNYDSLFIATIALASLIIRHKONI 189
DB 226 LKEVGTQTPWVLIFFITL--IKHAGNI 250

RESULT 15
US-09-778-927A-79
; Sequence 79, Application US/09778927A
; Patent No. US2002008342A1
; GENERAL INFORMATION:
; APPLICANT: KHOSRAVI, Rami et al.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
; TITLE OF INVENTION: VARIANTS OF ALTERNATIVE SPLICING
; FILE REFERENCE: 2786-0160P
; CURRENT APPLICATION NUMBER: US/09/778, 927A
; CURRENT FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: IL 134453
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: IL135341
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 79
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(434)
; OTHER INFORMATION: xaa = any amino acid, unknown, or other
US-09-778-927A-79

Query Match 9.9%; Score 105.5; DB 10; Length 434;
Best Local Similarity 25.0%; Pred. No. 0.0088;
Matches 53; Conservative 33; Mismatches 73; Indels 53; Gaps 12;
QY 11 YLIGSIPSGL-----WIG--QVFOINLREHSGNGTGTNTFRILG-----KKAGMA 55
DB 5 WLLGDDPAGSAGTGPWLSREQVFLISA---ASVNLGSMCMYSILGPPFPKAEKKGAS 61
QY 56 TFVIDFFKGTATLPIIFHLOGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIF 114
DB 62 NTIIGMIFGCA-----LFEL--LASLVFENLVHIGAKFM---FVAGMEVSGGVITLF 110
QY 115 GF-----APIF---CLYL-----AIIFFGALYLGSMISLSSVTASIAAIG----- 152
DB 111 GVLDKRPDGPVFVTFAMFLVRVMDAVSFAAAMTASSILAKAFNNVATVLSLETSGLG 170
QY 153 -VLLFPLFGFILSNYDSLFIATIALASLIIR 183

Db 171 LILGPPVGGFLYQSGYEVFPFIVLGCVVLLMV 202
Search completed: December 26, 2002, 02:58:43
Job time : 90.1872 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:27:20 : Search time 151.618 Seconds
(without alignments)
58.268 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITIVLLILAVLGSIPSL.....NKTENLVPGLNLTHODPKK 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	1071	100.0	213	1	Y851_STRPN	Q54916 streptococ
2	468	43.7	193	1	YNES_BACSU	Q45064 bacillus su
3	332	31.0	192	1	Y676_AQUAE	Q66905 aquifex aeo
4	300.5	28.1	222	1	YJ73_SYNV3	P73933 synechocyst
5	263.5	24.6	205	1	YGIH_ECOLI	P31056 escherichia
6	244.5	22.8	199	1	YGIH_HAEIN	P44603 haemophilus
7	223.5	20.9	239	1	Y247_MYCPG	P47489 mycoplasma
8	214.5	20.0	239	1	Y247_MYCPN	P75428 mycoplasma
9	213	19.9	220	1	YF09_HELPJ	Q921b1 helicobacte
10	211	19.7	220	1	YF09_HELPJ	Q921b1 helicobacte
11	117	10.9	304	1	YTRL_BUCSC	Q26039 buchneria ap
12	102	9.5	348	1	NU2M_CARAU	Q44601 buchneria ap
13	97.5	9.1	330	1	Y374_METJA	Q78680 carassius a
14	95.5	8.9	173	1	NU6M_LATCH	Q57819 methanococ
15	95.5	8.9	336	1	MRAY_STRPY	O03175 latimeria c
16	95.5	8.9	337	1	YE71_HAEIN	O99yk2 streptococ
17	95	8.9	402	1	YBJJ_ECOLI	O57130 haemophilus
18	95	8.9	532	1	YC18_HAEIN	P75810 escherichia
19	94.5	8.8	414	1	Y0AB_BACSU	Q57251 haemophilus
20	93	8.7	598	1	Y0AB_BACSU	Q45064 bacillus su
21	92.5	8.6	458	1	NUON_RICCN	P94532 bacillus su
22	92.5	8.6	492	1	YD19_METJA	Q92hh5 rickettsia
23	91.5	8.5	412	1	Y663_METH	O58715 methanococ
24	91.5	8.5	438	1	PBUX_BACSU	O26759 methanobact
25	91.5	8.5	552	1	CYHR_CANMA	P42086 bacillus su
26	91	8.5	497	1	DTPT_LACLA	P32071 candida mal
27	90.5	8.5	172	1	NU6M_BRARE	Q9m1x9 brachydanio
28	90.5	8.5	220	1	YQJA_ECOLI	P42614 escherichia
29	90.5	8.5	442	1	CITN_LACLA	P21608 lactococcus
30	90.5	8.5	452	1	YDJE_ECOLI	P38055 escherichia
31	90	8.4	276	1	NSOY_FSEST	P19845 pseudomonas
32	89.5	8.4	423	1	YHJV_ECOLI	P37660 escherichia
33	89.5	8.4	440	1	CAPE_STAAU	P39854 staphylococ

34	89	8.3	201	1	PSS_METJA	Q58609 methanococ
35	89	8.3	473	1	MOT4_CHICK	P57788 gallus gall
36	89	8.3	622	1	COX1_BACSU	P24010 bacillus su
37	88.5	8.3	170	1	NUOJ_BUCAI	P57260 buchneria ap
38	88.5	8.3	428	1	DTA_ECO57	Q8x5m2 escherichia
39	88.5	8.3	428	1	DTA_ECOLI	P37312 escherichia
40	88.5	8.3	439	1	IDNT_ECOLI	P39344 escherichia
41	88.5	8.3	546	1	YAO5_SCHPO	Q10084 schizosacch
42	88	8.2	162	1	Y159_PYRHO	O57898 pyrococcus
43	88	8.2	413	1	YGI8_AQUAE	O67545 aquifex aeo
44	88	8.2	450	1	GNUT_PSEAE	Q9a1j1 pseudomonas
45	88	8.2	450	1	VGLM_HSVB	P28948 equine herp

ALIGNMENTS

RESULT 1
Y851_STRPN
ID Y851_STRPN STANDARD; PRT; 213 AA.
AC Q54916;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein SP0851.
GN SP0851.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=7785;
RX MEDLINE=96313246; PubMed=8763932;
RA Pan X., Fisher M.;
RT "Cloning and characterization of the parC and parE genes of
Streptococcus pneumoniae encoding DNA topoisomerase IV: role in
fluoroquinolone resistance."
RT J. Bacteriol. 178:4060-4069(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
*Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.*
RL Science 293:498-506(2001).
CC -!- SUPCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z67739; CAA91549.1; -
CC TIGR; SP0851; -
CC InterPro; IPR003811; DUF205.
CC Pfam; PF02660; DUF205; 1.
CC TIGRFAMs; TIGR00023; DUF205; 1.
CC Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 2 22
CC POTENTIAL.

```
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT CONFLICT 168 188 S -> F (IN REF. 1).
SQ SEQUENCE 213 AA; 22928 MW; 22CB089C17750818 CRC64;

Query Match 100.08; Score 1071; DB 1; Length 213;
Best Local Similarity 100.08; Pred. No. 3.4e-74;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MITIVLLIAYLGGSPSLGWIGVFFQINLREHSGNGTGTNTFRILGKAGMATFVID 60
Db 1 MITIVLLIAYLGGSPSLGWIGVFFQINLREHSGNGTGTNTFRILGKAGMATFVID 60

Qy 61 FFKGTLATLLPIIFHLQGVSPFLIFGLAVIIGHTFFIFAGKGGKAVATSGVIFGAPIF 120
Db 61 FFKGTLATLLPIIFHLQGVSPFLIFGLAVIIGHTFFIFAGKGGKAVATSGVIFGAPIF 120

Qy 121 CLYLAIIFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDSLFTAILALASL 180
Db 121 CLYLAIIFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDSLFTAILALASL 180

Qy 181 IIRHKDNIAIKNKNTENLPVGLNLTHTODPKK 213
Db 181 IIRHKDNIAIKNKNTENLPVGLNLTHTODPKK 213

RESULT 2
YNES_BACSU STANDARD; PRT: 193 AA.
AC Q45064;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yneS.
GN YNES.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degree region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter."
RL Microbiology 142:3097-3101(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcher S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasehara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maul C., Medigue C.,
RA Medone D., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Neelam D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
```

```
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipatt A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73234; CAA97604.1; -
DR EMBL; Z99113; CAB13690.1; -
DR Subtilist; BG11826; yneS.
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 54 74 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
SQ SEQUENCE 193 AA; 20966 MW; C75803C399B97292 CRC64;

Query Match 43.78; Score 468; DB 1; Length 193;
Best Local Similarity 48.08; Pred. No. 7.9e-29;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

Qy 1 MITIVLLIAYLGGSPSLGWIGVFFQINLREHSGNGTGTNTFRILGKAGMATFVID 60
Db 1 MLJALLIAYLIGSPSLGVKLAKGIDIREHSGNLGATNAFTLGVKAGSVVINGD 60

Qy 61 FFKGTLATLLPIIFHLQGVSPFLIFGLAVIIGHTFFIFAGKGGKAVATSGVIFGAPIF 120
Db 61 ILKGTALATLPLFMHVD-IHPLLAGVAVLGHVFPFIFAKFKGKAVATSGVLLFYAPLL 119

Qy 121 CLYLAIIFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDSLFTAILALASL 180
Db 120 FITMVAVVFIFLYLTKFVSLSSMLTGIVTI-----YSFFV--HDTYLLIVVTLTIF 170

Qy 181 IIRHKDNIAIKNKNTENLPV 202
Db 171 VIYHRANKRIINKTEPKVKW 192

RESULT 3
Y676_AQUAE STANDARD; PRT: 192 AA.
AC O66905;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_676.
GN AQ_676.
OS Aquifex aeolicus
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
```



```

RA MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujoy M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT *The complete genome of the hyperthermophilic bacterium Aquifex
RT aolicus.*;
RL Nature 392:353-358(1998).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000701; AAC06869.1; -
CC InterPro: IPR003811; DUF205.
CC Pfam: PF02660; DUF205. 1.
CC TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
SQ SEQUENCE 192 AA; 20940 MW; EAD53C4016D63C00 CRC64;

Query Match 31.0%; Score 332; DB 1; Length 192;
Best Local Similarity 39.8%; Pred. No. 1.4e-18;
Matches 80; Conservative 39; Mismatches 68; Indels 14; Gaps 6;

QY 1 MITIVLLILAYLLGSLGWVFFQINLRHGSGMTGTNTPTIRILGKKGAMATFVID 60
DB 1 MKALFLVIFAYLLGSITTEGVIKLV-KGVDLNRVSGNMGVATNTVTRALGKKGVLVFFLD 59

QY 61 FFKGTALVLLPI-IFHLOGVSLPIFGLAVVIGHTFTPIFAGKGGKAVATSGVIFGFAP 119
DB 60 FFKGTALVIAVKGIDSWLTFGLASVLGHMYPVFFGKGGKGVATLGVVFAVSPS 119

QY 120 FCLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNY--DSLFTAIILAL 177
DB 120 VALFSFLVWGLGIFLWKRYVLSATITISAF-----LFLFV-AGYVNVLFMAIV--I 169

QY 178 ASLIIIRHKDNIARTKNKTN 198
DB 170 CALIIVRHRENINRLTGREH 190

RESULT 4
YJ73_SYNY3 STANDARD; PRT; 222 AA.
ID YJ73_SYNY3
AC P73933;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein sll1973.
GN SLL1973.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT *Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

```

```

RT entire genome and assignment of potential protein-coding regions.*;
RL DNA Res. 3:109-136(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90910; BAA17999.1; -
CC InterPro: IPR003811; DUF205.
CC Pfam: PF02660; DUF205; 1.
CC TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 56 76 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
SQ SEQUENCE 222 AA; 23534 MW; 4718AB2C20833360 CRC64;

Query Match 28.18; Score 300.5; DB 1; Length 222;
Best Local Similarity 34.2%; Pred. No. 3.7e-16;
Matches 75; Conservative 37; Mismatches 66; Indels 41; Gaps 6;

QY 1 MITIVLLILAYLLGSLGWVFFQINLRHGSGMTGTNTPTIRILGKKGAMATFVID 60
DB 6 LLCICLLLIITVLMGSIPTGYLAGLLCIDIREHGSKSTGATNVRTLCKPAAIAVLAD 65

QY 61 FFKGTALVLLPIIFHLOGVSLPI-----FGLLAVVIGHTFTPIFAGKGGKAVAT 109
DB 66 ISKGVMAVALVRAIYSGDWLPALPAQWNLTLGVAVLVGHGSKSIFLKFSGGKSVAT 125

QY 110 AGVIFGFAPICLYLAI-----IFFGALYLGSMISLSSVTASIAAVIGVLL-----FPL 158
DB 126 LGVLF-----MLNIMVLAAGTLATLFTVIFTRIVSLSSVAIAI-AVNGIALALQLPPPYLA 180

QY 159 FGFILSNYDSLFIATILALASLIIRHKDNIARTKNKTE 197
DB 161 FTFLAGMY-----VIVRHRTNIERILQGT 205

RESULT 5
YGIH_ECOLI STANDARD; PRT; 205 AA.
ID YGIH_ECOLI
AC P31056;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ygiH.
GN YGIH OR B3059 OR 24412 OR ECS3942.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT *The complete genome sequence of Escherichia coli K-12.*;
RN Science 277:1453-1474(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

```



```

QY 61 FFKGTLATLPI-IPHLOGVSLIFGLAV---IGHTEPIFAGKGGKAVATAGVIFGF 116
Db 63 MKUG-----MIPWAGYIIGLTOFELGVALGACUGHIFPIFOFGKGGVATAGAI--- 115
QY 117 APIFLYLAIIFFGALYGLSMISLSSVTASIAAVIGVLFLPLFGFILSNYDSLFIATILA 176
Db 116 APISNAVAGSGFTWIF-----VFLVSGYSSLSAVISALLVPFYVMWPKPEFTPEVALV-- 169
QY 177 LASLIIRHKONARIKNKNTENLVPW 202
Db 170 -CCLLIYRHNDIORLWRGOEDKV-W 193

RESULT 7
Y247_MYCGE
ID Y247_MYCGE STANDARD; PRT; 239 AA.
AC P47489;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG247.
GN MG247.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39703; AAC71467.1; -
DR TIGR: MG247; -
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
SQ SEQUENCE 239 AA; 27489 MW; D78CE976DEF621FD CRC64;

Query Match 20.9%; Score 223.5; DB 1; Length 239;
Best Local Similarity 31.8%; Pred. No. 2.5e-10;
Matches 74; Conservative 33; Mismatches 87; Indels 39; Gaps 10;

QY 1 MITIVLLIAYLLGSPSGMWIGVFFOI---NLREHSGNGTGTNTFRILCKKAGMATF 57
Db 9 ILVIFSLASGYLGST-----IFADIFSKILKNVRFSGSKNPGATNSRVFLKIGFLVA 64
QY 58 VIDFFKGLTATLPIIT---FHLQGV-----SPLIFGLL-----AVIGHTFTPIFAGFKGG 103
Db 65 IFDAFKGFFAFLTLTWLFRFGLQGYLTKEVYQSTFVLSLSCFAATIGHIPFLYKFKGG 124

QY 61 FFKGTLATLPI-IPHLOGVSLIFGLAV---IGHTEPIFAGKGGKAVATAGVIFGF 116
Db 63 MKUG-----MIPWAGYIIGLTOFELGVALGACUGHIFPIFOFGKGGVATAGAI--- 115
QY 117 APIFLYLAIIFFGALYGLSMISLSSVTASIAAVIGVLFLPLFGFILSNYDSLFIATILA 176
Db 116 APISNAVAGSGFTWIF-----VFLVSGYSSLSAVISALLVPFYVMWPKPEFTPEVALV-- 169
QY 177 LASLIIRHKONARIKNKNTENLVPW 202
Db 170 -CCLLIYRHNDIORLWRGOEDKV-W 193

```

```

QY 104 KAVATSAGVIFGFAIFCLYLAIIFFGALYGLSMISLSS-VTASIAAVIGVLFLPLFGFI 162
Db 125 KAIATTTGGSLLAISLWMLFICLLIWMITLTKVYSLASLTFVFLAVI--ILIPWLDYL 182
QY 163 -LSNYDSL-----FIATILA-----LASLIIRHKONARIKNKNTENLV 200
Db 183 YFFNSDPLKSYQNEWYIILFFCLWYWPVTVVVFWLHRANIIRLHCKESKI 235

RESULT 8
Y247_MYCPN
ID Y247_MYCPN STANDARD; PRT; 239 AA.
AC P75428;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG247 homolog (H91_orf239).
GN MPN350 OR MP486.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000047; AAB96134.1; -
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
SQ SEQUENCE 239 AA; 27439 MW; 6D4110A8253C9EBB CRC64;

Query Match 20.0%; Score 214.5; DB 1; Length 239;
Best Local Similarity 29.2%; Pred. No. 1.2e-09;
Matches 69; Conservative 40; Mismatches 82; Indels 45; Gaps 10;

QY 1 MITIVLLIAYLLGSPSGMWIGVFFOI---NLREHSGNGTGTNTFRILCKKAGMATF 57
Db 9 LLIVFSLVIGLYMGSV----MFADYFGKILNKDVRKLSKNPGATNSIRVFLKIGFLVG 64
QY 58 VIDFFKGLTIA---TLLPIIFHLQGVSL-----IFGLLAVIGHTFTPIFAGFKGG 103
Db 65 LCDALKGLFLAFVSEFLSFESFWMQLQVLYLVNVOYKVYLYVLSLSCFAATIGHIPFLYKFKGG 124
QY 104 KAVATSAGVIFGFAIFCLYLAIIFFGALYGLSMISLSS-VTASIAAVIGVLFLPLF 159
Db 125 KAIATTTGGSLLAISLWMLFICLLIWMITLTKVYSLASLTFVFLAVI--ILVPL 179
QY 160 GPI-----LSNYDSLFIATILA-----LASLIIRHKONARIKNKNTENLV 200

```



```
YTR1_BUCSC
ID YTR1_BUCSC STANDARD; PRT; 304 AA.
AC Q44601;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical transport protein in tpa3'region.
DE Hypothetical transport protein in tpa3'region.
OS Buchnera aphidicola (subsp. Schlechtendalia chinensis).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95261545; PubMed=7742976;
RA Lai C.-Y., Baumann P., Moran N.A.;
RT "Genetics of the tryptophan biosynthetic pathway of the prokaryotic
RT endosymbiont (Buchnera) of the aphid Schlechtendalia chinensis.";
RL Insect Mol. Biol. 4:47-59(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE EAMA TRANSPORTER FAMILY. STRONG, TO
CC S.TYPHIMURIUM PAGO.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U09185; AAA92793.1;
CC InterPro: IPR000620; DUF6.
CC Pfam: PF00892; DUF6; 2.
CC
CC Hypothetical protein; transport; Transmembrane.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 268 288 POTENTIAL.
SQ SEQUENCE 304 AA; 35051 MW; A6676E9610E55EAE CRC64;

Query Match 10.9%; Score 117; DB 1; Length 304;
Best Local Similarity 27.4%; Pred. No. 0.032;
Matches 57; Conservative 34; Mismatches 89; Indels 28; Gaps 11;

QY 1 MITVLLIAYLLGSLGPSLWGVFFQINLREHSGNGTGTNTFRILCKKAGMATFVID 60
DB 1 MIRKTIILLFIIIVSI---TW-GTTFIAIRIA-----SDTTPPICITGMRELLASFFLI 50
QY 61 FFKGTALTLPIIFHLQVSPILFGLLAVIGH-TPPIFAGKGRV-ATSAGVIFGAP 118
DB 51 FL-CFYTKPLF---PSNKIFQLIICIFYESLPFLILYGGRVNSTIASVFIMP 105
QY 119 IFCLYLALIFFG-ALYLSMSLSSVTSIAAIVI---GVLLFPLFGFILSNYDSLFATIL 175
DB 106 IIVFLSFFFNKLYFFOIFGL-----VLAIFLSIILFKREIELGDEKTINGVIALLL 159
QY 176 ALAS--LIIRKDNRIKNTENLVP 201
DB 160 AMTSHAIYLYSKERYSNISILTFNALP 187

RESULT 12
NU2M_CARAU
ID NU2M_CARAU STANDARD; PRT; 348 AA.
AC Q78680;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OF ND2.
OS Carassius auratus (Goldfish).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=P23 / Langsdorff; TISSUE=Oocyte;
RA Murakami M., Yamashita Y., Fujitani H.;
RT "The complete sequence of mitochondrial genome from a gynogenetic
RT triploid 'gimbuna' (Carassius auratus langsdorff).";
RL Zool. Sci. 15:335-337(1998).
CC -!- CATABOLIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE006953; BAA31239.1;
CC InterPro: IPR003917; NADHub_oxred2.
CC DR InterPro: IPR001750; Oxidored_q1.
CC DR Pfam: PF00361; oxidored_q1; 1.
CC DR PRINTS: PR01436; NADHDGNASE2.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 348 AA; 37879 MW; 0747C30C282486EB CRC64;

Query Match 9.5%; Score 102; DB 1; Length 348;
Best Local Similarity 23.9%; Pred. No. 0.48;
Matches 60; Conservative 35; Mismatches 84; Indels 72; Gaps 10;

QY 3 TIVLLIAYLLGSLGPSLWGVFFQINLREHSGNGTGTNTFRILCKKAGMATFVIDFF 62
DB 96 TMEFTALAKIGLAPMFWMEVQLQDL-----TGLILSTWQKLAPFAL--- 141
QY 63 KGLTALTLPIIFHLQVSPILFGLLAVIGH-TPPIFAGKGRV-ATSAGVIFGAP 110
DB 142 -----IIQTAQNIIDPL---LTLGLVSTLWGGWGLNQTQLRKILAYSSIAHMGW 189
QY 111 -GVIFGAP---IFCLYLALIFFGALYLSMSLSS-----VTASIAAVIG 152
DB 190 MIIVIQYAPQLTLLALGYIIMTSAFLTKLMSLTWKVSTLATTWSKSPILTATLVL 249
QY 153 VL--LPFLGFI-----LSNYDSLFATILALASLIIR-----HKDNRIKNT 196
DB 250 SLGLPLPTLGFMPKWLILQELTKQDLPITATTMALAALISLYFLRLCYLMTLISPNT 309
QY 197 ENLVPGLNLT 207
DB 310 NSTTPMRTOTT 320

RESULT 13
Y374_METJA
ID Y374_METJA STANDARD; PRT; 330 AA.
AC Q57819;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0374.
GN MJ0374.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
```

```

OX NCBI_TaxID-2190;
RW [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst H.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii*.
RR Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII MJ0201 AND TO H.INFLUENZAE AND
CC H.SONNUS HI0703.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensel@isb-sib.ch).
CC -----
CC EMBL: U67490; AAB98363.1; .
CC TRIGR: MJ0374; .
CC DR InterPro: IPR000252; DedA.
CC DR InterPro: IPR000326; PA_PTPase.
CC DR Pfam: PF00597; DedA; 1.
CC DR Pfam: PF01569; PAP2; 1.
CC DR SMART: SM00014; acidPPC; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 330 AA; 37018 MW; BA84B0170694C097 CRC64;
CC -----
Query Match 9.1%; Score 97.5; DB 1; Length 330;
Best Local Similarity 23.0%; Pred. No. 1;
Matches 47; Conservative 29; Mismatches 43; Indels 85; Gaps 9;
OY 57 FVIDPFKGLATLLPIFIHQ-----GVSPILFGLLAVIGHTF-----PIFA 98
DB :||| :||| :||| :||| :||| :||| :||| :|||
DB 26 FILSFEAFIQIPDPDVFVITGASFFGLNPITSAIVATIGTTLGGFLGYFLGDKLGHPIFI 85
OY 99 GPKKGRAV-----ATSAGVIFGFAP-----IPL-----122
DB ||| :||| :||| :||| :||| :||| :||| :|||
DB 86 KLEGEYLHKGEFFNKYGVGVIAGFSPLPYKVIAWLSGIFEMHKLFTVTGTIIRLP 145
OY 123 -YLAIFFGALYGLSMISSTASTAAVIGLVLLFPFLFGFILSNYDSLFTAI-----173
DB :||| :||| :||| :||| :||| :||| :||| :|||
DB 146 RLFAVAYFGDV-LGNINRLSDIN-----IYLFYLNSHNYIFDAIMPIISKTA 193
OY 174 --ILALASLIITRHKDNIARIKNK 195
DB ::| :||| :||| :||| :||| :||| :||| :|||
DB 194 YPLIAITSLIIF-----IKNR 209
RESULT 14
NU6M_LATCH STANDARD; PRT; 173 AA.
AC 003175;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH+ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
GN MTND6 OR ND6 OR NADH6.
OS Latimeria chalumnae (Latimeria) (Coelacanth).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Coelacanthiformes; Coelacanthidae; Latimeria.
OX NCBI_TaxID=7897;
RN [1]

```


Result No.	Score	Query		Length	DB	id	Description
		Match					
1	716.5	66.9	213	16	Q9CGW4		Q9cgw4 lactococcus
2	716	66.9	213	16	Q9A070		Q9a070 streptococc
3	621	58.0	168	2	Q9X972		Q9x972 streptococc
4	463.5	43.3	198	16	Q8YJ73		Q8y7j3 listeria mo
5	459.5	42.9	198	16	Q92C68		Q92c68 listeria in
6	441.5	41.2	202	16	Q99UC5		Q99uc5 staphylococ
7	385	35.9	194	16	Q8RRY9		Q8rry9 fusobacteri
8	311.5	29.1	198	16	Q8R9J2		Q8r9j2 thermoaena
9	298	27.8	206	16	Q9KCD3		Q9kcd3 bacillus ha
10	291.5	27.2	198	16	Q9RSV1		Q9rsv1 deinococcus
11	286.5	26.8	195	16	Q9M84		Q9m84 rhizobium l
12	282.5	26.4	205	16	Q8YC64		Q8yc64 brucella me
13	282	26.3	205	16	Q8URF1		Q8urf1 agrobacteri
14	275.5	25.7	198	16	Q97G69		Q97g69 clostridium
15	275.5	25.7	200	16	Q9JUL4		Q9jul4 neisseria m
16	274.5	25.6	199	2	Q9ZAF3		Q9zaf3 thermus the


```
Db 61 DLLKCTLATLPLFFHNGVSPILFGLLAVIGHTFISIFDRFKGKAVATSAGVILGFSPL 120
QY 120 FCLYLAIIFGALYLGSMISLSSVTASTAAVIGVLLPFLGFLSNYSDFIAIILALAS 179
Db 121 FLIYLLVVFIIIVLWLFMSISLSSVIGAVFALLGILFISIGFILTSDLLSIIIFVLAI 180
QY 180 LIIIRHKDNIARIKNNKNTLVPMGLNTHQDPK 212
Db 181 IILRHRTNLRKNNKNCESLVPFGLNLSKQEK 213

RESULT 2
Q9A070 PRELIMINARY: PRT; 213 AA.
ID Q9A070 AC
AC Q9A070:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein SPY0908.
GN SPY0908.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE065539; AAK33824.1; -
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 213 AA; 23369 MW; 6A9881232A09766A CRC64;

Query Match 66.9%; Score 716; DB 16; Length 213;
Best Local Similarity 64.8%; Pred. No. 3.9e-50;
Matches 138; Conservative 27; Mismatches 48; Indels 0; Gaps 0;

QY 1 MITIVLLIAYLLGSPGLWIGVFFQINLREHSGNLTGTTNTRILGKAGMATFVID 60
Db 1 MKLLFITAYLLGSPGLWIGVFFQINLREHSGNLTGTTNTRILGKAGMATFVID 60

QY 61 PFKGLTATLPIIFHQLQVSPILFGLLAVIGHTFIFAGFKGKAVATSAGVIFGAPIF 120
Db 61 MFKGTLISLLPIIFGNTSISSIAIGFAVLGHTFIFANFKGKAVATSAGVLLGFAPLY 120

QY 121 CLYLAIIFGALYLGSMISLSSVTASTAAVIGVLLPFLGFLSNYSDFIAIILALASL 180
Db 121 LFLFASIFVLVLFMSISLSSVASTAAVIGVLSVLPFAIHFLPNYDFLTIVILLAFI 180

QY 181 IIRHKDNIARIKNNKNTLVPMGLNTHQDPK 213
Db 181 IIRHKDNIARIKNNKNTLVPMGLNLSKQVPRK 213

RESULT 3
Q9X972 PRELIMINARY: PRT; 168 AA.
ID Q9X972 AC
AC Q9X972:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 17.9 kDa protein (Fragment).
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
```

```
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHI;
RA Vriesema A.J., Dankert J., Zaat S.A.;
RT "Isolation and characterization of promoter regions from Streptococcus gordonii CHI.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ236899; CAB40549.1; -
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
KW Hypothetical protein.
FT NON_TER 168 168
SQ SEQUENCE 168 AA; 17921 MW; A07262BD799A478A CRC64;

Query Match 58.0%; Score 621; DB 2; Length 168;
Best Local Similarity 72.1%; Pred. No. 1.3e-42;
Matches 119; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 3 TIVLLIAYLLGSPGLWIGVFFQINLREHSGNLTGTTNTRILGKAGMATFVIDFF 62
Db 4 TILGLIAYLLGSPGLWIGVFFQINLREHSGNLTGTTNTRILGKAGMATFVIDFF 63

QY 63 KGTATLPLPIIFHQLQVSPILFGLLAVIGHTFIFAGFKGKAVATSAGVIFGAPIFCL 122
Db 64 KGTATLPLPLHNGISPMIFGLIAGVLTFFIFABFKGKAVATSAGVYVXGSPFFS 123

QY 123 YLAIIFFGALYLGSMISLSSVTASTAAVIGVLLPFLGFLSNYSYD 167
Db 124 YLIITFIVTLGSMISLSSVAVAGFAIISVLPIPLGIIPLPSYD 168

RESULT 4
Q8Y7J3 PRELIMINARY: PRT; 198 AA.
ID Q8Y7J3 AC
AC Q8Y7J3:
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo1284.
GN lmo1284.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ESD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefl J., Kuhn M., Kunst F., Kurapat G.,
RA Mauguero E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591978; CAC99362.1; -
DR ListList; LMO1284; -
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21599 MW; 7807B5406DF05CD1 CRC64;

Query Match 43.3%; Score 463.5; DB 16; Length 198;
Best Local Similarity 47.8%; Pred. No. 7e-30;
Matches 96; Conservative 38; Mismatches 56; Indels 11; Gaps 4;
```

Db 176 VIWRHRANITIRNGEPEKIKW 197

RESULT 6

Q99UC5 PRELIMINARY; PRT; 202 AA.

AC Q99UC5

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical protein SAV1353.

GN SAV1353 OR Sali187.

OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and

OC Staphylococcus aureus (strain N315).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Staphylococcus.

OX NCBI_TaxID=158878, 158879;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);

RX MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-O., Ito T.,

RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,

RA Mizutani-U I Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,

RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,

RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;

RT "Whole genome sequencing of methicillin-resistant Staphylococcus

RT aureus";

RL Lancet. 357:1225-1240(2001).

DR EMBL; AF003362; BAB57515.1;

DR EMBL; AF003133; BAB42445.1;

DR InterPro; IPR003811; DUF205.

DR Pfam; PF02660; DUF205; 1.

DR TIGRFAMS; TIGR00023; DUF205; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 202 AA; 22322 MW; A9DA126B5731749C CRC64;

Query Match 41.2%; Score 441.5; DB 16; Length 202;

Best Local Similarity 42.9%; Pred. No. 4.2e-28;

Matches 90; Conservative 43; Mismatches 60; Indels 17; Gaps

Qy 1 MIVTVLLIAYLGSPGWTGQVFPQNLREHSGNGTGTNTPTFLGKKKAGMATFVID 60

Db 1 MMTIVMLLLSYLGAPPSPGVIGKLPFKKIDRFQFGSGNTGATNSRVLGRPAGFLVTFLD 60

Qy 61 FFKGTATLPIIFHVGVSPP-----LIFGLLAIVGHTPEIFAGFKGGRAVATSAGV 112

Db 61 IFKGFIVTFPMLPVPVHAGDPISTFTNGLIIVGLFAILGHVTPVILKFGGKAATVATAGV 120

Qy 113 IFGFAPIFCLYLAIIFFGALYLGMSITSSVASTAAIVGILLPFLFGFLSNYSDLFTA 172

Db 121 VLGVNPIPLLILAIIFVILKIFKYVSLASIAVAICCVISL-----IIDIYILLVVS 173

Qy 173 IILALASLIIRHKDNIARKNTENLVPW 202

Db 174 FLVSI--ILIRHSNIARIFRGEPEKIKW 201

RESULT 7

Q8RFY9 PRELIMINARY; PRT; 194 AA.

AC Q8RFY9;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical membrane-spanning protein FN0537.

GN FN0537.

OS Fusobacterium nucleatum (subsp. nucleatum).

OC Bacteria; Fusobacteria; Fusobacterium.

OX NCBI_TaxID=76856;


```
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vanathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AE002039; AAF11571.1;
DR TIGR: DR2021;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 20713 MW; 114362BD2E712BC1 CRC64;

Query Match 27.2%; Score 291.5; DB 16; Length 198;
Best Local Similarity 34.5%; Pred. No. 4.6e-16;
Matches 68; Conservative 39; Mismatches 81; Indels 9; Gaps 3;

OY 1 MITIVLLIAYLGSIPLGSLMTGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVID 60
Db 3 LTALLALLSYLIGAIIPAAMALARA-RGVDIRKVGSGNGATNRLSLGKGPALLVASPD 61

OY 61 FPKGTLATLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFKGKAVATSGVIFGFAPIF 120
Db 62 ILKGVLAVALRALGSAEWAALCGVLAVIGHNFSPLAFRGKGKGVATSGFVIAILDVPL 121

OY 121 CLVLAIFFGALYGMISLSVVTASIAAVIGVLLFPLFGFILSNVDSLFIAIILALASL 180
Db 122 GLTFTVLAATACHWLRFVSAGSIMGAFIAGALVLP-----RPTWDR---AAVLFALAL 173

OY 181 IIRHKDNIARIKNKTE 197
Db 174 LVWQHRENIKLGAGTE 190

RESULT 11
OY Q98M84 PRELIMINARY; PRT; 195 AA.
AC Q98M84;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein mlr0688.
GN MLR0688.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iodesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AF002995; BAB48229.1;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.

DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 20237 MW; 06FA01BE9DC7BEB9 CRC64;

Query Match 26.8%; Score 286.5; DB 16; Length 195;
Best Local Similarity 36.9%; Pred. No. 1.1e-15;
Matches 73; Conservative 37; Mismatches 75; Indels 13; Gaps 4;

OY 4 IVLLIAYLGSIPLGSLMTGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVIDPFK 63
Db 6 ILALVFGVLSGSIPLGSLMTGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVIDPFK 65

OY 64 GTLATLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFKGKAVATSGVIFGFAPIFCY 123
Db 66 GTAAVL--IAGHFAPETAVWAGLGLFGLFPVWLGKGGKGVATYGLVGLIGLAWQVALI 123

OY 124 LAIIFFGALYGMISLSVVTASIAAVIGVLLFPLFGFILSNVDSLFATILALASLI- 182
Db 124 FAVIWLAMAFLFYSSLAALTAAV-----IVPIALYFLS---APOIATVLFVVMISIVF 173

OY 183 IIRHKDNIARIKNKTE 200
Db 174 IIRHKDNIARIKNKTE 191

RESULT 12
OY Q8YC64 PRELIMINARY; PRT; 201 AA.
AC Q8YC64;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN BMEI10668.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Seikov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009702; AAL53910.1;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Complete proteome.
SQ SEQUENCE 201 AA; 20507 MW; 105CA44587B84CA1 CRC64;

Query Match 26.4%; Score 282.5; DB 16; Length 201;
Best Local Similarity 36.6%; Pred. No. 2.5e-15;
Matches 71; Conservative 31; Mismatches 81; Indels 11; Gaps 3;

OY 7 LILAYLGSIPLGSLMTGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVIDPFK 66
Db 15 LIFGVLSGSIPLGSLMTGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVIDPFK 74

OY 67 ATLLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFKGKAVATSGVIFGFAPIFCYLAI 126
Db 75 AAL--IAAHFGQNAATAGAGFAGFAGHLPVWLGKGGKGVATYGLVGLIGLAWAGALVFAA 132

OY 127 IFFGALYGMISLSVVTASIAAVIGVLLFPLFGFILSNVDSLFATILALASLIIRHK 186
Db 133 AWIVTALLARYSSLSALVASL--VVPIALY-----SRGNOALALFAIMTVIIFIKHR 183
```

```

QY 187 DNIRIKKNTENLV 200
DB 184 ANISRLNGTESKI 197

RESULT 13
Q8UFU1
ID Q8UFU1 PRELIMINARY; PRT; 205 AA.
AC Q8UFU1
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Atul306.
GN ATU1306 OR AGF_C_2402.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A. PubMed=11743193;
RX MEDLINE=21608551; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009092; AAL42312.1; ALT_INIT.
DR EMBL; AE008058; AAK87097.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 21678 MW; 973A5024E0CFDE46 CRC64;

Query Match 26.3%; Score 282; DB 16; Length 205;
Best Local Similarity 35.1%; Pred. No. 2.8e-15;
Matches 68; Conservative 40; Mismatches 74; Indels 12; Gaps 4;

QY 8 ILAYLLGSLPGWIGQVFFQINLRHSGNGTGTNTFRILGKAGMATFVIDFKGTLA 67
DB 19 LIGYLLGSLPGWIGWLRGKIGSGNIGATNVLRTGNKLAATLLDALKGTAA 78

QY 68 TLLPIIFHLOGV-SPLIFGLLAVTIGHTTPIFAGFGKGVKAVATSGAVIFGFAPICLYLAI 126
DB 79 VL--VANALWGEASLVAGFFAFGLHLPVWLGFGKGVAVYIGVLGAAPLMLAFAL 136

QY 127 IFFGALYLGSMISLVSTASIAIVGLFPLFGFILSNYDSLFIATILALASLIIRHK 186
DB 137 IWLATAFITRYSSLS-----ALLAMLIIPVALWLGPEKTAMLVTLSSVISN--WKHR 187

QY 187 DNIRIKKNTENLV 200
DB 188 ENIRLMAGTESRI 201

RESULT 14

```

```

Q97G69
ID Q97G69 PRELIMINARY; PRT; 198 AA.
AC Q97G69;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Uncharacterized conserved membrane protein, YgiH/Upp0078
DE family.
GN CAC2500.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Soucaille P., Daly M.J.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Smith D.R.;
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007749; AAK80454.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
KW Complete proteome.
SQ SEQUENCE 198 AA; 21099 MW; E50FEBDAFF82BBBA CRC64;

Query Match 25.7%; Score 275.5; DB 16; Length 198;
Best Local Similarity 32.0%; Pred. No. 8.9e-15;
Matches 65; Conservative 39; Mismatches 82; Indels 19; Gaps 4;

QY 1 MITVLLILAYLLGSLPGWIGQVFFQINLRHSGNGTGTNTFRILGKAGMATFVID 60
DB 2 ILVITVWVSLGCSIPGTGLITKLSGIDVTRKSGNIGSTNVKRVAGTKISMITQMD 61

QY 61 FFKGTATLPLPIIFHLOGVSP-----IFGLLAVTIGHTTPIFAGFGKGVKAVATSGAVI 113
DB 62 ILKGIIPVLLCMLIASIKLPISTSMVLSIIIVIAVLGHDTYTPFGFNGKGVNTTGAF 121

QY 114 FGFAPIFCLYLAIFFGALYLGSMISLVSTASIAIVG--VLFPPLFGFILSNYDSLFI 171
DB 122 FLAPAAVLAVAGAVYFVLRFTKIVSKIAVGTIMPTACIALRLPI-----EITY 172

QY 172 AILALASLIIRHKDNIRIKKNTKE 197
DB 173 CALIA-CGLLIRHKDNIRLVNNEE 197

RESULT 15
Q9JUL4
ID Q9JUL4 PRELIMINARY; PRT; 200 AA.
AC Q9JUL4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN NMA1261.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;

```

RT *Complete DNA sequence of a serogroup A strain of Neisseria

meningitidis 22491.

RL Nature 404:502-506(2000).

DR EMBL; AL162755; CAB84516.1; -

DR InterPro; IPR003811; DUF205.

DR Pfam; PF02660; DUF205; 1.

DR TIGRFAMs; TIGR00023; DUF205; 1.

KW Complete proteome.

SO SEQUENCE 200 AA; 20796 MW; DF1BEE95E87C43FF CRC64;

Query Match 25.7%; Score 275.5; DB 16; Length 200;

Best Local Similarity 34.1%; Pred. No. 9e-15;

Matches 70; Conservative 39; Mismatches 82; Indels 15; Gaps 4;

Oy 1 MITIVLLILAYLGSIFSGLWIGVFFQINLRHGSGNTGTTNTRILGKKAGMATFVID 60

Db 1 MNIPAVAVSYLIGLSFAVIVSKYIGMDPRTYGSNGPGATNVLRSCKKKAALTLIGD 60

Oy 61 FFKGTATLPIIFH---LQGVSPILIFGLLAVIGHTPIFAGFKGKAVATSAGVIFGF 116

Db 61 AAKGLVAVLLARVLOEPLGLSDSAIAAVALAALVGHWPVFFGFKGKGATALGVLLAL 120

Oy 117 APIFCLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLF-PLFGFILSNYDSLFIAIL 175

Db 121 SPTTALVCALILWYMAFGKVKVSSLAULTATIAAPLAALFFMPHTSWIFAT-----L 171

Oy 176 ALASLIIRHKDNIAR-IKKNTENL 199

Db 172 AAILVLLRHKSNIILNLIKESKI 196

Search completed: December 26, 2002, 01:16:19
Job time : 689.791 secs

CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 60/000,788
APPLICATION NUMBER: 03-JUL-1996
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0511
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SFO ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 438 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 115..144
OTHER INFORMATION: /note= "Encompasses TM 4 of Figure"
OTHER INFORMATION: 4
FEATURE:
NAME/KEY: Region
LOCATION: 323..357
OTHER INFORMATION: /note= "Encompasses TM 9 of Figure"
OTHER INFORMATION: 4
FEATURE:
NAME/KEY: Region
LOCATION: 359..386
OTHER INFORMATION: /note= "Encompasses TM 10 of Figure"
OTHER INFORMATION: 4
US-08-677-049-9

Query Match 8.5%; Score 91.5; DB 2; Length 438;

Best Local Similarity 24.9%; Pred. No. 0.048;

Matches 44; Conservative 31; Mismatches 51; Indels 51; Gaps 11;

QY 47 ILGKAGMA---TFV---IDFFKGTATLLPI---IFHLOGVSLIFGLLAVIGHTF-- 94
DB 29 IVKRAMGLTEQTLVLSIDIFMCGVATLLQVNSRFFGI-----GLPVLGCTFTA 80
QY 95 --PIFAGFGKAVATSGAVIFGPAIFCLYLAIIFFGALYLGSMISL--SSVTASIAAV 150
DB 81 VSPMI-----AIGSEGVSTVYGISIASGILVILI-SFFFGKLVSPFPVVTGVSVTI 132
QY 151 IGVLLPPL-----RGFILSNYDSLFIATLILALASLIIRHKDNATARKN 194
DB 133 IGITLMPVAMNNMAGGSADEFG-DLSN-----LALAFVLSIIVLLYRFTKGFIS 183

RESULT 9

US-09-134-001C-4744

Sequence 4744, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4744

LENGTH: 518

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4744

Query Match 8.5%; Score 91.5; DB 4; Length 518;

Best Local Similarity 23.9%; Pred. No. 0.06;

Matches 39; Conservative 33; Mismatches 72; Indels 19; Gaps 7;

QY 46 RIUGKAGMATFVIDFFKGTATLLPIIFHLOGVSLIFGLLAVIGHTFPFAGF-KGOK 104
DB 117 RVYTEAGDAITLPDFKRLDDKNII-----KIISGLIIVVFYTLTHSGFVSGK 169
QY 105 AVATSGAVIF--GFAPIFCLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLPFLGFI 162
DB 170 LFESAGLNYHAGLLIVAIIVIFYTFFGG-YLA--VSITDFQGVIMLIAMVWVPIVALL 226

QY 163 -LSNYSLSFIAIILALASLIIRHKDNATARKNTENLVPMGL 204
DB 227 KLNGWDTFHDIAQMKPTNLDLFRGTTVLGIV-----SLFSNGL 264

RESULT 10

US-09-134-001C-2912

Sequence 2912, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 2912

LENGTH: 400

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-2912

Query Match 8.4%; Score 89.5; DB 4; Length 400;

Best Local Similarity 21.8%; Pred. No. 0.071;

Matches 57; Conservative 35; Mismatches 60; Indels 109; Gaps 12;

QY 1 MITIVLLILAYLVSGIPSGMLTQGVFFOINLREHSGMGTGTTNTPR-----ILGKKA 52
DB 166 MWVFLGIVSLVLGLI-----VLVFMKE-----NTNTIINKEEKMKNRNLNKKF 211
QY 53 GMATFVIDFFKG-----TLATLLPIIFHLOGV-----SPLIFGLLA-- 88
DB 212 VIFFSIAVFFEGAGYIITGTFLVAIVKSPISVAEYATLSWMFVGLGAIPSTLIWSLIAEK 271
QY 89 -----VIGHTEPIFAGFGKAVATSGAVIFGPAIFCLYLAIIFFGALYL 134
DB 272 ISYKAIYAGAFILQIISVCLPVT-----HEIFSLVISSVLFGGTFL 313
QY 135 G-----SMISLSSVTASIAAVIGVLLPFLFGFIL-----SNYD-SL 169
DB 314 GLTTLFISKQSLMYKTDDBPLNLSVLTIVYS---LGQMLAPMFAGILIGKSNYNIAL 369
QY 170 FTATILALASLI--IIRKDN 188
DB 370 IFATVLLILGLISTIFS YKQN 390

RESULT 11

US-09-134-001C-3501

Sequence 3501, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 3501

LENGTH: 434

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3501

Query Match 8.2%; Score 88; DB 4; Length 454;
Best Local Similarity 23.3%; Pred. No. 0.12;
Matches 60; Conservative 33; Mismatches 81; Indels 84; Gaps 13;
Qy 13 LGSIPSLWTI--GOVFOINLREHSGNTGTTNFRILGKKGAMATFVIDF-----FKGTLL 66
Db 15 IGSMLGFLFGAGNLFPHLGGTAGSNWTAN---LG-----FLTAIGLPLGII 63
Qy 67 A---TLPIIFHLOQVSLIFGLLAVIGHTF---PIPA-----GFKGKAVATAGVI 113
Db 64 AIGVSKTNGVFEISSRSKIYGLFTGLVIGPFPALPRLATTSFEIAFSPFISSGTA 123
Qy 114 FGPAIFCLYLAIFGALYLSMISLSSVTASIAAVIGVLLPLF----- 159
Db 124 QALLPIF---SILFFGAVL-----FSRKPSKILDYIGKFLNPVFLILGIVVVLAFIR 174
Qy 160 -----GFILSNYDSLFAITAILALASLII-----IRHKDNIAIR 191
Db 175 PMGISHAPVSADYSNVLKGFIDGYNLTDALASLAFGIIIVTIKKLGTNPNTIAK 233
Qy 192 --INKTENLVPMLNIT 207
Db 234 ETLKSGTISIAMGVIT 251

RESULT 12

US-09-165-396-4

Sequence 4, Application US/09165396

Patent No. 6441134

GENERAL INFORMATION:

APPLICANT: BECKER, JEFFREY W.

APPLICANT: LUBKOWITZ, MARK A.

TITLE OF INVENTION: ISOLATED CANDIDA ALBICANS OLIGOPEPTIDE TRANSPORTER GENE

FILE REFERENCE: 372.6520P

CURRENT APPLICATION NUMBER: US/09/165,396

CURRENT FILING DATE: 1998-10-02

EARLIER APPLICATION NUMBER: PCT/US98/02332

EARLIER FILING DATE: 1998-02-06

EARLIER APPLICATION NUMBER: 60/037,859

EARLIER FILING DATE: 1997-02-07

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4

LENGTH: 799

TYPE: PRT

ORGANISM: S. cerevisiae

US-09-165-396-4

Query Match 8.2%; Score 87.5; DB 4; Length 799;
Best Local Similarity 21.9%; Pred. No. 0.3;
Matches 46; Conservative 37; Mismatches 82; Indels 45; Gaps 9;

Qy 4 IVLLIAYLLGSLPSGLWIGOVFOINLREHSGNTGTTNFRILGKKGAMATFVIDFFK 63
Db 279 LIVLIGSLFIWVPGFLTGLSYFNVILGSKTRHNTANT--IFGTQSLGA----- 329
Qy 64 GTLATLLPIIFHLOQVSLIFGLLAVIGHTFPIFAGFKGKAVATAGVIFGPAIF-CL 122
Db 330 -----LPITFDYQVSQAMSG--SVFATPFVVSAN-----TYASVLIFFVIVLPCL 373
Qy 123 YLAIIFFGALYL-----GSMISLSSVTASIAAVIG-----VLLPFLGFLILSN 165
Db 374 YFTNTWY-AKYPVITSGSYDNTQNKYNTKILNEDYSINLEKYKEYSPVFPFSLYS- 431
Qy 166 YDSLFIATILALASLIIIRHKDNIAIRKNK 195
Db 432 YALNFAAVIAVVFVHCILTHGKDIVAKFDR 461

RESULT 13

US-09-245-808-1

Sequence 1, Application US/09245808

Patent No. 6313277

GENERAL INFORMATION:

APPLICANT: Doyle, L. Austin

APPLICANT: Abruzzo, Lynne V.

APPLICANT: Ross, Douglas D.

TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

FILE REFERENCE: encodes it

CURRENT APPLICATION NUMBER: US/09/245,808

CURRENT FILING DATE: 1999-02-05

EARLIER APPLICATION NUMBER: 60/073763

EARLIER FILING DATE: 1998-02-05

NUMBER OF SEQ ID NOS: 7

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 655

TYPE: PRT

ORGANISM: Human MCF-7/AdrVp cells

US-09-245-808-1

Query Match 8.0%; Score 86; DB 4; Length 655;
Best Local Similarity 19.2%; Pred. No. 0.34;
Matches 47; Conservative 43; Mismatches 75; Indels 80; Gaps 12;

Qy 9 LAYLLGSLPSGLWIGOVFOINLREHSGNTG-----TNTFRILGKKGAMATFVID-- 60
Db 396 IAOIIVTVVLGLVIGAIYFGLKNDSTGIONRAGVLFLLTN--QCFSSVSASVELFVVEKK 453
Qy 61 -----PFKGTPL-----ATLLP-----IIFHLOQVSP-----LIFGL 86
Db 454 LFIHEIYSYRVSSVFLGKLLSDLLPMTMLPSIIFTCIVYFMLGKPKADAFVMMFTL 513
Qy 87 LAVIGHTFPFAGFKGKAVATAGVIFGPAIFCLYLAIIFFGALYLSMISLSSVTAS 146
Db 514 MVAISASSMALAIAAGSVSVATLLM-----TICFVFMIFSGLL-----VNLTTI-AS 563
Qy 147 IAAVIGVLLPFLGFLSNDYSLFIAILALASLIIIRHKDNIAIRKNKTNENLVPMLGNL 206
Db 564 WLSWLOVFSIPRYGF-----TALQHNFEFG-----QNFCEP-GLNA 597
Qy 207 THODP 211
Db 598 TGNPN 602

RESULT 14

US-09-134-001C-5389

Sequence 5389, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOX

FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 5389

LENGTH: 901

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5389

Query Match 8.0%; Score 86; DB 4; Length 901;
Best Local Similarity 24.0%; Pred. No. 0.52;
Matches 52; Conservative 31; Mismatches 72; Indels 62; Gaps 10;

Qy 28 QINLREHSGNTGTTNFRILGKKGAMATFVIDFFKGTLL-ATLLPIIFHLOQVSLIFGL 86

```
Db 169 QIETOGGAMNSEPGTSEIVGIVARVILLITF--GSLIAGMPIISAIGLSSV-GI 225
Oy 87 LAVIGHTFPI-----PAGEKGG-----KAVATSAGVIFG 115
Db 226 IALTYIFDIPNFTLTAVMIGLAVGIDYSLFLFRFKELKKGVDVVEAIAVGTAGS 285
Oy 116 FAPIFCULAIIFGALYLGSMISLSSVTASIAAVIGVLLPFLPGFILSNYDSLFAIIL 175
Db 286 -----AVIFAGLTVMTAVCGLSLVGIDFLAVMG-----FASAISVLFVLAALTL 330
Oy 176 ALASLIIRHKNIARKNKNTNLPWGLNLTHODPK 212
Db 331 -LPALISIPHK--IKIKDKPK-----SKDKP 355
```

RESULT 15

```
US-09-134-001C-5000
; Sequence 5000, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5000
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5000
```

```
Query Match      8.0%; Score 85.5; DB 4; Length 190;
Best Local Similarity 22.9%; Pred. NO. 0.074;
Matches 47; Conservative 28; Mismatches 69; Indels 61; Gaps 9;

Oy 1 MITIVLLIAYLLGSLGWLIGQVFFQINLREHSGNLTGTTNFRILGKKGAGMATFVID 60
Db 19 LMTAICILG-LVPSVPLPFMPVPLVQ-----NIGIFLAGIILGRKLG----- 62
Oy 61 FFKGTATLLPIIFHLQGVSLIFGLLAVIGHTFPFIFAGKGGKAV--ATSAGVIFGPAP 118
Db 63 -----TSVIVFLLVATG--LPVLSGGRGGIGVFPAGPSAGFLFLY-P 101
Oy 119 IFCLYLAIIFGALYLGSMISLSSVTASIAAVIGVLLPFLPGFILSNY-----DSLFI- 171
Db 102 VVAVFIGIIRDAYLHKINFLVFIAT----LVIGVLGDLILGTLINGFIHPIPSKAFIL 157
Oy 172 -----AIIALASLIIRH 185
Db 158 SFTMPGDIIRAIATIASLIGAILNH 182
```

Search completed: December 26, 2002, 02:47:12
Job time : 134.781 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 24, 2002, 22:13:55 ; Search time 11247.8 Seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgatacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_be.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	642	100.0	6171	6	BD003717	BD003717 Polynucle
2	642	100.0	10624	1	AE007390	AE007390 Streptoco
3	640.4	99.8	6812	1	SPPARCETP	267739 Streptoco
4	640.4	99.8	10828	1	AE008451	AE008451 Streptoco
5	637.2	99.3	248254	2	SPNEU1903	AL44925 Streptoco
6	288.4	44.9	25021	1	AB028896	AB028896 Streptoco
7	275.6	42.9	11545	1	AE006539	AE006539 Streptoco
8	275.6	42.9	11573	1	AE010024	AE010024 Streptoco
9	274	42.7	53354	1	AE014149	AE014149 Streptoco
10	246.6	38.4	13211	1	AE006332	AE006332 Lactococc
11	244.6	38.1	771	1	SGO236899	AJ236899 Streptoco
12	153.4	23.9	250050	1	AL591978	AL591978 Listeria
13	147	22.9	7563	1	AF084044	AF084044 Listeria
14	124	19.3	195269	6	AX417035	AX417035 Sequence
15	124	19.3	349980	6	AX417044	AX417044 Sequence
16	103.8	16.2	9839	1	AE010565	AE010565 Fusobacte
17	102.4	16.0	3014	1	AF269800	AF269800 Staphyloc
18	102.4	16.0	3014	6	AX145118	AX145118 Sequence
19	102.4	16.0	4105	1	AF269437	AF269437 Staphyloc
20	102.4	16.0	4105	6	AX144757	AX144757 Sequence
21	100.8	15.7	295350	1	AP004826	AP004826 Staphyloc
22	100.8	15.7	303750	1	AP003133	AP003133 Staphyloc
23	100.8	15.7	346900	1	AP003362	AP003362 Staphyloc
24	100	15.6	6595	1	AF024713	AF024713 Bacillus
25	100	15.6	26170	1	BC170DEGR	273234 B.subtilis
26	100	15.6	233780	1	BSUB0010	299113 Bacillus su
27	96	15.0	594	6	AX144133	AX144133 Sequence
28	69.4	10.8	10530	1	AE013956	AE013956 Yersinia
29	68.2	10.6	11198	1	AE004138	AE004138 Vibrio ch
30	67.4	10.5	10177	1	U32713	U32713 Haemophilus
31	66	10.3	7218	6	I66494	I66494 Sequence 14
32	65.2	10.2	300550	1	AP001512	AP001512 Bacillus
33	61	9.5	1033	1	LLRXNRDEF	X92690 L.lactis nr
34	61	9.5	12043	1	AE009702	AE009702 Brucella
35	59	9.2	12422	1	AE004494	AE004494 Pseudomon
36	58.4	9.1	20604	1	AE008847	AE008847 Salmonell
37	57.9	9.0	258050	1	AL627278	AL627278 Salmonell
38	56.5	8.8	346897	1	AP002995	AP002995 Mesorhizo
39	56.4	8.8	14651	1	AE012788	AE012788 Chlorobiu
40	56.2	8.8	8634	1	AF281816	AF281816 Acholepla
41	55.2	8.6	10029	1	AE005535	AE005535 Escherich
42	55.2	8.6	307962	1	AP002564	AP002564 Escherich
43	54.8	8.5	15663	1	AE001796	AE001796 Thermotog
44	53.6	8.3	1641	1	ECOBACTRA	L12966 Escherichia
45	53.6	8.3	11307	1	AE000387	AE000387 Escherich

ALIGNMENTS

RESULT 1
BD003717
LOCUS BD003717 6171 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003717
VERSION BD003717.1 GI:18631678
KEYWORDS JP 2001501833-A/37.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 6171)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
Fannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 37 13-FEB-2001;

RESULT 2	AE007390/c	10624 bp	DNA	linear	BCT 31-AUG-2001
LOCUS	AE007390	10624 bp	DNA	linear	BCT 31-AUG-2001
DEFINITION	Streptococcus pneumoniae TIGR4 section 73 of 194 of the complete genome.				
ACCESSION	AE007390	AE005672			
VERSION	AE007390.1	GI:14972313			
KEYWORDS	Streptococcus pneumoniae TIGR4.				
SOURCE	Streptococcus pneumoniae TIGR4				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
REFERENCE	1 (bases 1 to 10624)				
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouiri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Fraser, C.M.				
TITLE	Complete genome sequence of a virulent isolate of Streptococcus pneumoniae				
JOURNAL	Science 293 (5529), 498-506 (2001)				
MEDLINE	21357209				
PUBMED	11463916				
REFERENCE	2 (bases 1 to 10624)				
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouiri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Fraser, C.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
FEATURES	Location/Qualifiers				
source	1..10624				
gene	/organism="Streptococcus pneumoniae TIGR4"				
CDS	/strain="TIGR4"				
	/db_xref="taxon:170187"				
	126..362				
	/gene="SP0840"				
	126..362				
	/gene="SP0840"				
	/note="identified by Glimmer2; putative"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAK74971.1"				
	/db_xref="GI:14972314"				
	/translation="MKILRYLELCFILSPALPFIKGTNADNGRCFVETVYGFLLM EHAIVTAVFICSLIAFLKNDGRNLLRVVIAP"				
	736..1326				
	/gene="SP0841"				
	736..1326				
	/gene="SP0841"				
	/note="identified by Glimmer2; putative"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAK74972.1"				
	/db_xref="GI:14972315"				
	/translation="MSKMYAENPDAAHDIHELRYDLGLGKMTFLTDAGVFSKRMVDV GSQLLCLDVGQETLDVGGCGPLGLSLVKYVQATVQ				

RESULT 2
AE007390/c
LOCUS
DEFINITION
Streptococcus pneumoniae TIGR4 section 73 of 194 of the complete genome.
VERSION
AE007390.1 GI:14972313
KEYWORDS
ORGANISM
Streptococcus pneumoniae TIGR4.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE
1 (bases 1 to 10624)
Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
Complete genome sequence of a virulent isolate of Streptococcus pneumoniae
Science 293 (5529), 498-506 (2001)
MEDLINE
21357209
PUBMED
11463916
REFERENCE
2 (bases 1 to 10624)
Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., and Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
Direct Submission
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..10624
/organism="Streptococcus pneumoniae TIGR4"
/strain="TIGR4"
/db_xref="taxon:170187"
126..362
/gene="SP0840"
126..362
/gene="SP0840"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74971.1"
/db_xref="GI:14972314"
/translation="MKILKRYILELCFILSFALPFIKGTNADNGRCFVETYYGFTFLM
EHAIVTAVFICSEFLIAFLKNDGRNGLLRVIAF"
736..1326
/gene="SP0841"
736..1326
/gene="SP0841"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK74972.1"
/db_xref="GI:14972315"
/translation="MSKMYAENPDAAHDHIELRVLLGKMTFLTDAGVFSKKMYDF
GSOLLKCLDYNQGVYDVGCGYGLSLVKVYGVQATMVDINRALDAQRNAEK
NNAKATFQSNIVARYQGRFDHVISNPPIRAGQVHVHIEILKSKDFLEIGDGLTIVIQ
KKQAPSAKSKMEDVFGNCEILAKDKGYILRSVKE"
1323..2600
/gene="SP0842"

KPHRVLAGAUAEOGFYGDFFSTVEDVNDOTLLVQSNKGTIYESRLQDLNLSRTSNG

SFISDITSDREVDAYLQEVVTEDEK*

BASE COUNT 2106 a 1318 c 1555 g 1833 t

ORIGIN

Query Match 99.8%; Score 640.4; DB 1; Length 6812;
Best Local Similarity 99.8%; Pred. No. 1.7e-150;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGATTACAATAGTATTTTAAATCTAGCTATCTGCTGGGTTCCATTCATCTGCTC 60
DB 1119 ATGATTACAATAGTATTTTAAATCTAGCTATCTGCTGGGTTCCATTCATCTGCTC 1060
QY 61 TCGATTGGACAAGTATCTTCAATCAATCTACGCGACATGTTCTGGTAACACTGGA 120
DB 1059 TCGATTGGACAAGTATCTTCAATCAATCTACGCGACATGTTCTGGTAACACTGGA 1000
QY 121 AGCACAACACCTTCGCAATTTAGTAAAGAGCTGGTATGGCAACCTTTCTGATTGAC 180
DB 999 AGCACAACACCTTCGCAATTTAGTAAAGAGCTGGTATGGCAACCTTTCTGATTGAC 940
QY 181 TTTTCAAGGAACCTTACCAAGCTGCTTCCGATTTATTTTCATCTACAAGCGTTTCT 240
DB 939 TTTTCAAGGAACCTTACCAAGCTGCTTCCGATTTATTTTCATCTACAAGCGTTTCT 880
QY 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTCCCTATCTTTTCAGGATT 300
DB 879 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTCCCTATCTTTTCAGGATT 820
QY 301 AAAGTGGTAAGCTTCCCAACCACTGCTGGAGTGATTTTCGGATTTTCGGCTATCTTTC 360
DB 819 AAAGTGGTAAGCTTCCCAACCACTGCTGGAGTGATTTTCGGATTTTCGGCTATCTTTC 760
QY 361 TGCTCTACCTTGGCAATATCTTCTTTGGAGCTCTCTATCTTGGCAGTAGATTTTCAC 420
DB 759 TGCTCTACCTTGGCAATATCTTCTTTGGAGCTCTCTATCTTGGCAGTAGATTTTCAC 700
QY 421 TCTAGTGTACAGCATCGATTGGGCTGTTATCGGGTCTGCTCTTCCACTTTTGGT 480
DB 699 TCTAGTGTACAGCATCGATTGGGCTGTTATCGGGTCTGCTCTTCCACTTTTGGT 640
QY 481 TTTATCTCTGAGTAACATGACTCTCTCTTCATCGCTTATTTATCTTAGCACTTCTGCTAGTT 540
DB 639 TTTATCTCTGAGTAACATGACTCTCTCTTCATCGCTTATTTATCTTAGCACTTCTGCTAGTT 580
QY 541 ATTATCATCTGCAATGACATATAGCTGATCAATAAATAAATAAATAAATTGGTC 600
DB 579 ATTATCATCTGCAATGACATATAGCTGATCAATAAATAAATAAATAAATTGGTC 520
QY 601 CCTTGGGATTTGAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCT 642
DB 519 CCTTGGGATTTGAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCT 478

RESULT 4
AE008451/c 10828 bp DNA linear BCT 13-SEP-2001
DEFINITION Streptococcus pneumoniae R6 section 67 of 184 of the complete genome.
ACCESSION AE008451
VERSION AE007317
KEYWORDS AE008451.1 GI:15458341
SOURCE Streptococcus pneumoniae R6.
ORGANISM Streptococcus pneumoniae R6
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1 (bases 1 to 10828)
AUTHORS Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.
Genome of the bacterium Streptococcus pneumoniae strain R6
J. Bacteriol. 183 (19), 5709-5717 (2001)
21429245
11544234
2 (bases 1 to 10828)
Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
Location/Qualifiers
1..10828
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
103..1638
/gene="ABC-NBD"
/note="spr0748"
103..1638
/gene="ABC-NBD"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0748"
/product="ABC transporter ATP-binding protein - ribose/galactose transport"
/protein_id="AAK99552.1"
/db_xref="GI:15458342"
/translation="MAHENVEMRDITKVFQGVANDKINLHRLKKEIHALLENGAG KSLMNLKAGLEPTSGEIVAGNVNLDSPKASUGIGHVQHFQMLVETVAENI ILSGLTGNLVDIAGASKIERSGLAVDSKADISVAGQFVETLKYLRG ADILFDEPTAVLTPSEIDELMAIMKLVKREKSIILITHRLDEIVASVDRVIRG KSIETVEATGATNADLAEMVGRSVFKTEKQAKPEVLSIKDLVNNRNGVPVK NLSLDVRAGEIVGAGIDNGOSELIOATIGLRKVESGIELKGSIVGLHPRITEL SVGHVPEDRHREDGLILENMISENIALOTYYKPKSKNGILNYSITSAKKLMEEFDV RAASLVPAAALSGNQAKIAREIDRPOLLIVSQPTRGDLVGAIEYIHKRLIEER DNGKAVLVVSFELDEILNVSDRIAVIHDGKIQGVISVPTTNKQGLGLVMAGNLGK SDV"

FEATURES
source

gene
CDS
1604..2689
/gene="ABC-MSP"
/note="spr0749"
1604..2689
/gene="ABC-MSP"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0749"
/product="ABC transporter membrane-spanning permease - ribose/galactose transport"
/protein_id="AAK99553.1"
/db_xref="GI:15458343"
/translation="METWERRRVMKSKLQOISVPLISVFLGILLGAIWVIFGYDAIW GYEELFYAFSGSLRGIGEIFRANGPLVLIGLGFVAVASRAGFNVGLPQALACHILSC WFAUSHPDMPLMILATIVIALIAGGIVGALPGILRAYLGTSEVIVITMMNIIIVY GNATFHAFKDFMQSDSTIRVANATYQTPMLTGLTNSRNIGIFFAITAVAIWF MLKKTTLGFEIRAVGNLPHASEYAGISAKRTIILSMIISGALAGLGAGVIGTQNY VVOGSSLAIFNGMAVSLAANSPIGILFAAFLEGLVGVAGPMNAAOVPSLVSIVT ASIIFVSVHYLLIERFVAPKKQVKGK"
2692..3648
/gene="ABC-MSP"
/note="spr0750"
2692..3648
/gene="ABC-MSP"

gene
CDS

/note="Putative"
/codon_start=1
/transl_table=11
/label_spr0750
/product="ABC transporter membrane-spanning permease -
ribose/galactose transport"
/protein_id="AAK99554.1"
/db_xref="GI:15458344"
/translation="MSIITLLPLLVSSMLIYSAPLIPTSGGVFSESGVGVVNVVGLGEI
MVMGAFGVNFLEFAEQGAATPMLSVAGLVGSVFSIIHAAATYHFRADHVSGT
VNLMPALAVFLVKVYNGQTDNLSTQTFGRDFPVLANIPIVIGDIFFKTSLLGYL
AIAFSFLAFTQFGLRLRSVGEHPQADTLGINVYKRWYLGVIISGFLGIGCA
IYAGISVNSVTIVGPGFIALAAMIFGKNPIGAMLSLIFGLSOSLAVIGSOLPF
LQGPAYVLOIAPVITLVLAAFFGKAVAPKADGGINIKSK"
3733..3945
/gene="IS861-truncation"
/note="spr0751"
3733..3945
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0751
/product="Degenerate transposase (orf1)"
/protein_id="AAK99555.1"
/db_xref="GI:15458345"
/translation="MKLSYDDKVOIYELRKGQISLEKLSNKGFINNINRYMIKLIDR
YGEFVAKGNRYSPDLKQEMTHKV"
4081..4248
/gene="IS861-truncation"
/note="spr0752"
4081..4248
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0752
/product="Degenerate transposase (orf1)"
/protein_id="AAK99556.1"
/db_xref="GI:15458346"
/translation="MGRNPKTRPEERTELERLOAENEYLAENAILKKLRELRLKEK
EKERQKLFPK"
4248..4616
/gene="IS861-truncation"
/note="spr0753"
4248..4616
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0753
/product="Degenerate transposase (orf2)"
/protein_id="AAK99557.1"
/db_xref="GI:15458347"
/translation="MTEFSLDLLKAIKLARSTYYYHLKOLDKTDKDOELKAEIQSIF
IEHGNTAYRRIYLELRNGYLVNHRKRVQHLWKYSIYKLKRDRENILLIKETLARKQ
RISFKNLKALKOWNSATQM"
4592..5080
/gene="IS861-truncation"
/note="spr0754"
4592..5080
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0754
/product="Degenerate transposase (orf2)"
/protein_id="AAK99558.1"
/db_xref="GI:15458348"
/translation="MEOCYDVTVEFAIPVSTOKLYLSPVLDFGNSEIIAYNLSTSPNL
EOVOTMLEQAFTEKHVENTILHSDQGWYOHDSYHRELESKGIQASNRKGNSDNGM
MESFGLIKSEMFGYEKTFKSLNLEQALIDYIDYNNRRIKVKLGLSPVQIRTKS
FG"
complement(5123..5764)
/gene="spr0755"
complement(5123..5764)
/gene="spr0755"

/note="Putative lipoprotein precursor"
/codon_start=1
/transl_table=11
/label_spr0755
/product="Conserved hypothetical protein"
/protein_id="AAK99559.1"
/db_xref="GI:15458349"
/translation="MITIVLLIAYLLGSPGMLIGQVFFQINLREHSGNGTGTNT
FRILGKAGNATVIDFFKGLTLLPIIFHLQGVSPILFGLLAVIHTPIFAGFKG
KRAIVASAGVIFGFAPICLYLAIIFFGALYLSMISLSVTSIAAIVIGLVLPFLFG
FILNSDLSLFIALLASLIIRIHKNDIARIKNKTENLVPMLNLTHODPKK"
5900..7843
/gene="parE"
/note="spr0756"
5900..7843
/gene="parE"
/EC_number="5.99.1.1"
/codon_start=1
/transl_table=11
/label_spr0756
/product="Topoisomerase IV subunit B"
/protein_id="AAK99560.1"
/db_xref="GI:15458350"
/translation="MSKKEININYNDDAIOVLEGLDAVPRKPMVIGSTDGAGLHL
WIEIVNADVEALSGRDRIDVTINKDGLTVODHGRGMPTHMGAGIPTVEVITIL
HAGKFGQGGYKTSGLHGVSSVNALSWLEIVEITRDGAVLYKRFENGGRKPVTLK
KICTALKSKTGTVPMDATIFSTDFKYNITISERLNSAFLLKKNVTLSLTKRDTDE
AIEFHYENGQVDFVSYLNEDKEILTPVLYPEGEDNGFOVEALQYNDGFSNLSLVN
NVRTKDGTHETGLKSAITKVMNDYARKTGLLKEKDNLEGSYDREGLAAVLSILVPE
EHLQFEGOTKDKLGSPLARPVDGIADVKTFFLMENGELASNLIRKAIKARDAREAA
RKARDSRNGKNKKDKGLLSGLKLTPAOSKNPAKNELYLVEGDSAGSAGKQGRDRFQ
AIPLRGKVINTAKMADILKNEEINTMIYITGAGVADFSIEDANYDKIIIMTDAD
TDGAHOTLLTTFYRMRPLVEAGHYIALPPLYKMSKKGKKEVAVAWTDCLEEE
LRQFGKGATLQRYKGLGEMNADQLMWTETNPRTLIIRVTIEDLARAERRVNLMDG
KVEPRKVEDNVKFTLEATVE"
8254..10734
/gene="parC"
Query Match 99.8%; Score 640.4; DB 1; Length 10828;
Best Local Similarity 99.8%; Pred. No. 1.7e-150;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGATTACAATAGTATTATTAACTCCTAGCCCTATCTGCTGGTTCGATCCATCTGGTCTC 60
|||||
Db 5764 ATGATTACAATAGTATTATTAACTCCTAGCCCTATCTGCTGGTTCGATCCATCTGGTCTC 5705
Qy 61 TGGATTGGACAAGTATCTTTTCAATCAATCTACGCGAGCATGCTTCTGGTAAACACTGGA 120
|||||
Db 5704 TGGATTGGACAAGTATCTTTTCAATCAATCTACGCGAGCATGCTTCTGGTAAACACTGGA 5645
Qy 121 ACGACCAACACCTTCCCGCATTTTAGGTAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 180
|||||
Db 5644 ACGACCAACACCTTCCCGCATTTTAGGTAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 5585
Qy 181 TTTTTCAAAGGAACCCCTAGCAACGCTGCTTCCGATTTATTTTTCATCTACAGGGCTTTCT 240
|||||
Db 5584 TTTTTCAAAGGAACCCCTAGCAACGCTGCTTCCGATTTATTTTTCATCTACAGGGCTTTCT 5525
Qy 241 CCTCTCATCTTTGGACCTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
|||||
Db 5524 CCTCTCATCTTTGGACCTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 5465
Qy 301 AAAGGTGGTAAAGGCTGCGCAACAGCTGCTGGAGTATTTTCGGATTTTGGCCCTATCTTC 360
|||||
Db 5464 AAAGGTGGTAAAGGCTGCGCAACAGCTGCTGGAGTATTTTTCGGATTTTGGCCCTATCTTC 5405
Qy 361 TGTCTCTACCTTGGCATATCTCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACCTG 420
|||||
Db 5404 TGTCTCTACCTTGGCATATCTCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACCTG 5345
Qy 421 TCTAGTGTACACAGCATCGATTTCGGCTGTTATCGGGGTCTTATCGGGGTCTTGTCTTTTCCACTTTTGGT 480
|||||
Db 5344 TCTAGTGTACACAGCATCGATTTCGGCTGTTATCGGGGTCTTATCGGGGTCTTGTCTTTTCCACTTTTGGT 5285


```

Qy 481 TTTATCCTGAGTAATGACTCTCTTCATCGCTATTATCTTAGCATTGCTAGTTTG 540
    |||||||
Db 5284 TTTATCCTGAGTAATGACTCTCTTCATCGCTATTATCTTAGCATTGCTAGTTTG 5225

Qy 541 ATTATCATTCGTCATAAGGACAATATAGCTGATCAAAAATAAAACTGAAAAATTTGGTC 600
    |||||||
Db 5224 ATTATCATTCGTCATAAGGACAATATAGCTGATCAAAAATAAAACTGAAAAATTTGGTC 5165

Qy 601 CTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAAATAA 642
    |||||||
Db 5164 CTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAAATAA 5123

RESULT 5
SPNEU1903/c      248254 bp  DNA  linear  HTG 11-JUL-2001
LOCUS
DEFINITION      Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
in ordered pieces.
ACCESSION      AL449925
VERSION        AL449925.1 GI:11545150
KEYWORDS        HTG; HTGS_PHASE2.
SOURCE          Streptococcus pneumoniae.
ORGANISM        Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE      1 (bases 1 to 248254)
AUTHORS        Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
and Garcia-Bustos, J.F.
TITLE          Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
JOURNAL        Microb. Drug Resist. 7 (2), 99-125 (2001)
MEDLINE        21335329
PUBMED        11442348
REFERENCE      2 (bases 1 to 248254)
AUTHORS        Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
Garcia-Bustos, J.F.
TITLE          Direct Submission
JOURNAL        Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT        * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES        Location/Qualifiers
source          1..248254
                /organism="Streptococcus pneumoniae"
                /serotype="19f"
                /db_xref="taxon:1313"
                /clone="G54"
BASE COUNT      74105 a 49201 c 48476 g 76331 t 141 others
ORIGIN

Query Match      99.3%; Score 637.2; DB 2; Length 248254;
Best Local Similarity 99.5%; Pred. No. 8.7e-150;
Matches 639; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGATTACAATAGTTTATTAATCCTAGCCTATCTGCTGGGTTCGATTCGATTCGGTCTC 60
    |||||||
Db 2546 ATGATTACAATAGTTTATTAATCCTAGCCTATCTGCTGGGTTCGATTCGATTCGGTCTC 2487

Qy 61 TGGATTGGACAAGTATTTCTTTCAATCAATCTACGCGAGCATGGTTCGGTAACACTGGA 120
    |||||||
Db 2486 TGGATTGGACAAGTATTTCTTTCAATCAATCTACGCGAGCATGGTTCGGTAACACTGGA 2427

Qy 121 ACGACCAACACTCCCGCATTTTATAGTAAAGAAAGCTGATGGCAACCTTTGTGATTGAC 180
    |||||||
Db 2426 ACGACCAACACTCCCGCATTTTATAGTAAAGAAAGCTGATGGCAACCTTTGTGATTGAC 2367

```

```

Qy 181 TTTTCAAGGAACCCCTAGCAAGCGTCTGTTCCGATTATTTTTCATCTACAAAGCGTTTCT 240
    |||||||
Db 2366 TTTTCAAGGAACCCCTAGCAAGCGTCTGTTCCGATTATTTTTCATCTACAAAGCGTTTCT 2307

Qy 241 CCTCTCATCTTTGGACATTTTGGCTGTTATCGGCCATACCTTCCTATCTTTTGCAGGATTT 300
    |||||||
Db 2306 CCTCTCATCTTTGGACATTTTGGCTGTTATCGGCCATACCTTCCTATCTTTTGCAGGATTT 2247

Qy 301 AAAGTGTTGAAGCGTCTGCAACCAAGTCTGGAGTGAATTTTCGATTTTGGCCCTATCTTC 360
    |||||||
Db 2246 AAAGTGTTGAAGCGTCTGCAACCAAGTCTGGAGTGAATTTTCGATTTTGGCCCTATCTTC 2187

Qy 361 TGTCTCTACCTTTCGATTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACCTG 420
    |||||||
Db 2186 TGTCTCTACCTTTCGATTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACCTG 2127

Qy 421 TCTAGTCTCAGACATCGATTGGGCTGTTATCGGGTTCCTGCTCTTTTCCACATTTTGGT 480
    |||||||
Db 2126 TCTAGTCTCAGACATCGATTGGGCTGTTATCGGGTTCCTGCTCTTTTCCACATTTTGGT 2057

Qy 481 TTTATCCTGAGTACTATGACTCTCTCTTCATCGCTATTATCTTAGCATTGCTAGTTTG 540
    |||||||
Db 2066 TTTATCCTGAGTAACTATGACTCTCTCTTCATCGCTATTATCTTAGCATTGCTAGTTTG 2007

Qy 541 ATTATCATCTCTATAAGGACAATATAGCTCGTATCAAAAATAAAACTGAAAAATTTGGTC 600
    |||||||
Db 2006 ATTATCATCTCTATAAGGACAATATAGCTCGTATCAAAAATAAAACTGAAAAATTTGGTC 1947

Qy 601 CTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAAATAA 642
    |||||||
Db 1946 CTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAAATAA 1905

RESULT 6
AB028896/c      25021 bp  DNA  linear  BCT 15-AUG-2001
LOCUS
DEFINITION      Streptococcus agalactiae type-specific capsular polysaccharide
biosynthesis gene region (cpsIa5 to beat gene), complete and
partial cds.
ACCESSION      AB028896
VERSION        AB028896.2 GI:15186709
KEYWORDS        Ufacil DNA glycosylase; ung; orf1; NeuA; NeuB; NeuC; CpsIaI;
CpsIaK; CpsIaJ; CpsIaI; CpsIaH; CpsIaG; CpsIaF; CpsIaE; CpsIaD;
CpsIaC; CpsIaB; CpsIaA.
SOURCE          Streptococcus agalactiae DNA.
ORGANISM        Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE      1
AUTHORS        Yamamoto, S., Miyake, K., Koike, Y., Watanabe, M., Machida, Y., Ohta, M.
and Iijima, S.
TITLE          Molecular characterization of type-specific capsular polysaccharide
biosynthesis genes of Streptococcus agalactiae type Ia
JOURNAL        J. Bacteriol. 181 (17), 5176-5184 (1999)
MEDLINE        99395021
REFERENCE      2 (bases 1 to 25021)
AUTHORS        Miyake, K., Yamamoto, S., Koike, Y., Watanabe, M. and Iijima, S.
TITLE          Direct Submission
JOURNAL        Submitted (14-JUN-1999) Katsuhide Miyake, Nagoya University,
Department of Biotechnology, School of Engineering, Furo-cho,
Chikusa-ku, Nagoya, Aichi 464-8603, Japan
COMMENT        (E-mail:miyake@proc.nubio.nagoya-u.ac.jp, Tel:81-52-789-4278,
Fax:81-52-789-3221)
On Aug 14, 2001 this sequence version replaced gi:5381177.
FEATURES        Location/Qualifiers
source          1..25021
                /organism="Streptococcus agalactiae"
                /db_xref="taxon:1311"
gene            1..375
                /gene="cpsIa5"
                <1..375
                /gene="cpsIa5"
                /function="putative histidine protein kinase"

```

```

/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAB62877.1"
/db_xref="GI:15186710"
/translation="DHVDTWKFKELKESNHPVLAPSERHSDSLTSKSLAECLQLTD
LDEVKSTIIQLQKFEKSVKRNPLAHLIKPDEQLNTTFFSQAFLDQIIFLAKVIG
VEDVTNHFYDTWNKLLIKILE"
/gene="cpsiaa"
/complement(400..1323)
/gene="cpsiaa"
/complement(400..1323)
/gene="cpsiaa"
/function="LysR-type transcriptional regulator"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAB62878.1"
/db_xref="GI:15186711"
/translation="MRIQQLQYVKKIVETGSMNEAKQLYITOPSLSNVARNLETEMG
IQIRPKGITLTKDGEFLSTARQILEQTALLEERYKGDNTSRELFVSSSHAFV
VNAFVALFNGDVTQYELFRETNRTEIIDVKNRSEIGVLFNSYNRDLTKLFDD
NSLIATFTTTHIFVFSKSNPLANRKKLSMKDLEDYPYLSYDQGLHNSFYFSEEMMS
QIHPKISVSDRATFNLMLIGLDGYVATGILNSKLNKGDEIVAIPLDVDDVIDIVI
RHDKANLSKMGQKFDYLLLEVSFNHKNKOS"
/complement(1447..1452)
/evidence-experimental
1453..1457
/evidence-experimental
/complement(1470..1475)
/evidence-experimental
1478..1482
/evidence-experimental
/complement(1481..1486)
/evidence-experimental
/complement(1507..1512)
/evidence-experimental
1512..1519
/gene="cpsiaa"
1512..1519
/gene="cpsiaa"
/function="regulator"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAB82275.1"
/db_xref="GI:5381178"
/translation="MSNHSRQOKKHSHTPLRVINFLVIFILLVSVSLFMYRHHF
LAFHLAVIYGVVLVLIILASLELCINKKARIEFTIILVLASIFVATLGYKSTIDL
TNLNKTAISEIEMSVIVPKDSKININAEVSKLAAPVKNDTSNITDLIEHKSEGI
SITPKQTDSDAYNRINKGDSQAMVNNAYVSLIELSTPDFKSOIKTIYTIKKIKI
NRKNTNKEGVNIYISGIDTFGSIIVSRSDVINIMTNTNTHKVLTTTPRDAYVK
IPDGGNQYDKLTHAGLYGVETSMKTLNENLDYIARINFSSEFLKLDLGGVTV
YNDQAFPSKGNFDPVQVTLNSEALGFVRERYSLQGGDNDGRNOEKVIAIINK
LASSQSVTKLNSITSOLOQTSVQTNMTIDNLDLNNLSTGQRTVESQALTGHGVTG
ELPSYAMPQAQLYMSIDQSSLSNAKSKINTWEE"
2975..3706
/gene="cpsiaa"
2975..3706
/gene="cpsiaa"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAB82276.1"
/db_xref="GI:5381179"
/translation="MIDIHSHIVFDVDDGPKTLEESLSLEESVROGVRIIVTSRHR
KGMETPEDIIIFKNFSIVKHEAKRFEHLQILYGGELYTSDMLEKLLKLOIPTLNTT
KFALIESQTSKWDKHTALSNNVLMLGITPVVAHIERYNALENOKERKVELINMGCTT
QINSHLKOKLFDNKHFRKRRARYFLEENLVHFVASDMHNDLVRPFPFLAEAYKIC
RDFGKERANQLFIENQAQSIKNHYI"
3715..4407
/gene="cpsiaa"
3715..4407
/gene="cpsiaa"

```

```

/function="chain length determination (putative)"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAA82277.1"
/db_xref="GI:5381180"
/translation="MNKIANTVEINIFNLLKLLKKKKEKITFVAIAPATAGLYSLF
IYTPQYISSRIYVINPNNISITAOQLQAGSFANDYKEIITSDVLEKVISSEKL
NYPSSQLQKITVSIKLDTRVISISVEDANPKMSQRLANSVREAAVSKAVTQVEDI
TTLEKGNLKPAPSSPNIKRNVILGIFVAGLSTIVLVIIMGILDDRVTBEDIEKALGL
TSLGIVPDLNKL"
418..14922
/misc_feature
/feature="corresponds to DDBJ Acc#:AB017355"
4418..5107
/gene="cpsiaa"
4418..5107
/gene="cpsiaa"
/function="chain length determination (putative)"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAA82278.1"
/db_xref="GI:5381181"
/translation="MTRLEIVDSKLPQAKKTEYFNAIPTNIOFSCKENKILAITSVR
EKGKSTTSLSLSLAQAQFKTLTLLDADTRNSVMSGTFKATGTIKGLTNYLSGNADL
GDIIETNPRLMVPSGKVPNPPTALLQNAFYFNKMEIAKNIFDIITDTPPIGLV
DAAITANACDGFILVTQAGRIKRNVEKAKEQMSGSKFLGIILNKVSESATYGDY
GNYGKRDRKK"
5243..5248
/-35_signal
5265..5270
/-10_signal
5334..6509
/gene="cpsiaa"
5334..6509
/gene="cpsiaa"
/function="glucosyltransferase"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAA82279.1"
/db_xref="GI:5381182"
/translation="MVLKYSFYIIFSSLFFIFKNSFTTTLSPFSFTIAMSILLYL
LNSFLKYRYKSYAKFSRDTKVLITNKDSLKMTNRKYDHNHYIACVILDSSEKDCY
DLKNSLRINKDALTSCLTVDQAFINPIELFKYQIODIINDIEMGVIVNVI
VEALSPDNIGEKRIOTFEYSVITYSMKEYKSHLIAKRELDITGAIIGLIGCVI
FLYPOIRKGGPAIFSONVRNGRIFRYKPSRMVDAEQIKKDLLVHNMHTGLMFK
LDDPRITIKGFIKRTSIDELPQFYNVLKGDMSLVGTRPPTVDETERNSTOKRRLS
FKPGITGLMOISGRNNITDFDEIVLDVQYINEMSWSDIKIILLTLKVLVLLGTGAK"
6533..6982
/gene="cpsiaa"
6533..6982
/gene="cpsiaa"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAA82280.1"
/db_xref="GI:5381183"
/translation="MKICLVSGSGHLAHLNLLKPKGNKDRFWTFDKEDARSILRE
EIVYHCFPTNRRNVLNNTILAFVLRKRPDVIISSGAVAVPFFYIGLFGCKT
VIIEVDFDRDKPPTLTKLVYPVTDKFIQVOMEEMKKVYPKAINLGGIF"
6982..7524
/gene="cpsiaa"
6982..7524
/gene="cpsiaa"
/function="galactosyltransferase"
/codon_start-1
/transl_table-11
/product="cpsiaa"
/protein_id="BAA82281.1"
/db_xref="GI:5381184"
/translation="MIFVTVGTHQEQFNRLIKEVDRLKGTGATDQEVFIQTGYSDPEP

```

```
ONQWSEFLSYDDMSYMKAEIIVITHGGPATPMNAVSKGKKTIVVPROBQFGEHVN
HOYDFELKFLKTELDYILNISELENIKEKNISTSKVISONNDPCFSKNEHFINV
NKYILLEKKEIENISQIC"
7563..8702
/gene="cpsiaH"
7563..8702
/gene="cpsiaH"
/function="CP polymerase (putative)"
/codon_start=1
/transl_table=11
/product="cpsiaH"
/protein_id="BAA82282.1"
/db_xref="GI:5381185"
/transl_table="MOLLLALLIVLLICSSYKKMKFMYMAEIFFIVFVILVTSIL
LHSEFTPDPRILAAFNLSLIGIVSVALRWYKNTTLELDKILKAFENGLILFELG
GTYICLHNNIQNISIFGRDLGSDWINGMHTORAMGFFEYSNLIPMTVVTNIIYI

Query Match      44.9%  Score 288.4;  DB 1;  Length 25021;
Best Local Similarity 65.6%  Pred. No. 4.5e-62;
Matches 421;  Conservative 0;  Mismatches 221;  Indels 0;  Gaps 0;

Oy 1 ATGATTACATAGTTTATTAATCCTAGCCTATCTGCTGGGTTGCGATTCCATCTGGCTC 60
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 19280 ATGAATATTAAATATGATTATGATTCGCTATCTGTTAGGCTCTATCCAACTGGACTT 19221

Oy 61 TGGATTGACAAGTATCTTTCAATCAATCACTACGCGAGCATGCTTCTGGTAACACTGGA 120
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 19220 TGGATTGGTAATAATTTTACCAGTCAATCTTCGTAACACGGTAGTGGTAATACCTGGC 19161

Oy 121 ACGACCAACACTTCCGCACTTTTAGGTAAGAAAGCTGCTATGCGCAACCTTTGTGATTGAC 180
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 19160 ACCAGTAACCTTCCGATTTTAGGTCATAAAGCTGCGATTGTGCTGACTATTGAC 19101

Oy 181 TTTTCAAGAACCCCTAGCAACCGCTGCTCCGATTAATTTTCACTACAAAGCGTTTCT 240
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 19100 ATTTTAAAGGACTCTAGCAACACTTATCTCTATCATATTAGGTATAACACAGATATC 19041

Oy 241 CCTCATCTTTGACTTTTGGCTGTTATCGGCATACCTTCCCTATCTTTCGACGATTT 300
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 19040 CCAATTTTATCGGTTTTTGGTATTAATAGTGCACACTTTCCTATTTTTCACAATTC 18981

Oy 301 AAAGTGGAAGGCTGTCGCAACCAAGCTGCTGGAGTGAATTTTCGGATTTCGCCCTTCTTC 360
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 18980 AAAGTGGAAGGCTGTCGTCACAGCGCTGGGTTTTTTCGCTAGGATTTCGACCTTCTTT 18921

Oy 361 TGTCTACCTTGGATATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTGCTACTG 420
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 18920 TTCTTTTACTTATAGTATCTTTTATTAACGCTTTTATCTTTTATAGCATGATTTCCCTA 18861

Oy 421 TCTAGTGCACAGCATGATTGCGGCTGTTATCGGGTCTGCTCTTCCACTTTTGGT 480
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 18860 TCCAGTATTACGATGCTGCTAGGATTTCTTATGCTCTTAATTTTCCATTAGTAGGC 18801

Oy 481 TTTATCTGAGTAATATGATCTCTCTCTATCCTGCTATTAATCTTATAGCACTTGGCTAGT 540
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 18800 TTTATTAACGAGCTAGCACTGCGATATTTACCAGCTGCTGTTATCTTAATGCTCTCACG 18741

Oy 541 ATTATCTCTGATAGGACAATATAGCTGCTATCAAAATAAATAAATTTGGTCT 600
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 18740 ATTATATCCGACATAGGATAATATCAACAGTATTCGAAAGCAAGCAAAATTTGGTT 18681

Oy 601 CCTTGGGATTGAACCTAACCCATCAAGATCTCTAAAAATAA 642
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 18680 CCGTTTGGCTTAACCTTAAGTAACAAAAATAAATAAATAA 18639

RESULT 7
AE006539/c
LOCUS
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 68 of 167 of
the complete genome.
ACCESSION AE006539 AE004092
VERSION AE006539.1 GI:13622059
```

KEYWORDS
SOURCE
ORGANISM

Streptococcus pyogenes M1 GAS.
Streptococcus pyogenes M1 GAS.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
AUTHORS

1 (bases 1 to 11545)
Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Tuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
21192684
MEDLINE
11296296

REFERENCE
AUTHORS

2 (bases 1 to 11545)
Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Tuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Direct Submission

TITLE
JOURNAL

Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
Oklahoma City, OK 73104, USA
Location/Qualifiers

FEATURES
SOURCE

1..11545
/organism="Streptococcus pyogenes M1 GAS"

/strain="SF370"

/serotype="M1"

/db_xref="taxon:160490"

362..1075
/gene="deod2"

/note="SPY0894"

362..1075
/gene="deod2"

/note="Best Blastp hit = sp|Q56037|DEOD_STRTR PURINE
NUCLEOSIDE PHOSPHORYLASE (INOSINE PHOSPHORYLASE) (PNP)
>gi|1276873|gb|AAC44007.1| (U40830) Deod [Streptococcus
thermophilus] >gi|1588604|prf|12209356A deod gene
[Streptococcus thermophilus]"

/codon_start=1

/transl_table=11

/product="putative purine nucleoside phosphorylase"

/protein_id="AAK33813.1"

/db_xref="GI:13622060"

/translation="MSIHISAKKGDIAKILLPGLPRAKFIENFLEDAVCNEVRN
MFQGTGTHGRVSMGMPISIIYARELIVDYGKTLIRVGTAGADPEVHVRN
LVLAQAATNSIIRNDFPFPQIADEFLLDKAYHAREMGVTHVGNVLSDDVY
TNPNRMALGKGVKAIEMEAALYLAQHVKALGIMTISDNLNDPTEDTTABER
QTTFTDMKVGLETLIAND"
1068..1856
/gene="SPY0895"

/note="SPY0895"

1068..1856
/gene="SPY0895"

/note="Best Blastp hit = gb|AAB81305.1| (L47121) histidine
protein kinase [Carnobacterium piscicola]"

/codon_start=1

/transl_table=11

/product="histidine protein kinase"

/protein_id="AAK33814.1"

/db_xref="GI:13622061"

/translation="MTNQTLDLIDVAYNHAFAIKALPNIPKTAALYLEMKERR
ELNLAFAEAENRTIDYHCSLMNQSLDEQIANVILDEVKNGKAIIDFVRS
VSPILYRLRLITSEIPKFIYFTKNDQYDTHFQAMLESDEHVFYKAYLSOKSR
NVTTLADMLTSLPQIKDLVLRHFRKAVRNPLAHLIKPFDFEELHRTTHFFSS
QAFLENIITATFSGVYRREPFFDDNNAIKKLSLRQSTV"
complement(1935..2840)
/gene="cpsy"

/note="SPY0898"

complement(1935..2840)
/gene="cpsy"

/note="Best Blastp hit = emb|CAB36977.1| (Y17221) Cpsy
protein [Streptococcus agalactiae]"

/codon_start=1

gene
CDS

gene
CDS


```

Qy 301 AAAGTGGTAGGCTGTCGCAACCAAGTCTGGAGTGAATTTTCGAGTTTGGCCCTATCTTC 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11150 AAAGTGGTAGGCGGTAGCAACAAGTCTGGTGTATTTGCTAGGCTTTGCTCCGTTATAT 11091
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 TGTCTCTACCTTGCGATTATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11090 CTCCTTTTTTTAGCATTAICTTTGCTTTTAGTCTCTATTTATTTAGTATGATATCTTTA 11031
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 TCTAGTGCACAGCATGATTCGGCTGTTATCGGGGTTCTGCTTTTCCACTTTTTTGGT 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11030 GCTAGTGTGGTTTCAGCTATCGTTGGTGTATCTGTCTGTTTAAACATTTCTCGCCATCAT 10971
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 TTTATCTCGAGTAAGTACTGCTCTCTTCATCGCTATTTATCTTAGCAGTCTGCTAGTTTG 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10970 TTTCTTTTACCAAAATTAAGTACTCTTTTAACTTTTATTTGATATTAAGTCTGCGTTCA 10911
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 541 ATTATCATCTGTCATAAGGACAATATAGCTGTCATCAAAATAAATAAATGAAATTTGGTC 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10910 ATTATTAACACACAAGATAAATAGTCTGATCAACATCATCACTGAAATCTTAATA 10851
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 601 CTTGGGGTGAACCTTAACCCATCAAGATCCTCAAAAATAA 642
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10850 CTTGGGGCTAAATTTAAGCAACAAGTCTGCTCAAAAATAA 10809
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS AE010024/c 11573 bp DNA linear BCT 03-APR-2002
DEFINITION Streptococcus pyogenes strain MGAS8232, section 72 of 173 of the
complete genome.
ACCESSION AE010024 AE009949
VERSION AE010024.1 GI:19748120
KEYWORDS
SOURCE
ORGANISM Streptococcus pyogenes MGAS8232.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 11573)
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Genome sequence and comparative microarray analysis of serotype M18
group A Streptococcus strains associated with acute rheumatic fever
outbreaks
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)
21927593
11917108
REFERENCE 2 (bases 1 to 11573)
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Direct Submission
Submitted (31-JAN-2002) Laboratory of Human Bacterial
Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St.,
Hamilton, MT 59840, USA
FEATURES
source
1. .11573
Location/Qualifiers
/organism="Streptococcus pyogenes MGAS8232"
/strain="MGAS8232"
/db_xref="taxon:186103"
362. .1075
/gene="deob"
/gene="spyM18_0955"
362. .1075
/gene="deob"
/notes="best blastp match gb|AAK33813.1| (AE006539)
putative purine nucleoside phosphorylase [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11

```

```

/product="putative purine nucleoside phosphorylase"
/protein_id="AAL97596.1"
/db_xref="GI:19748121"
/translation="MSIHSAKKGDIADKILLPGDPLRAKFAENFLEDAVCFNEVRN
MFGYTGTYGHRVSVMTGMPGISIYARELIVDYGVKTLIRVGTAGIDPEVHVRE
LVLAOAAATNSNIIRNDFEFPDFOIADFLGLDKAVHIAREMGVTHVGNLSSDVFY
TNMPERNMAGKLGKVKAIEMEAALYVLAQHHVKALGIMTISDNLNDPTEDTTAEER
QTFTDMKVGLETLLAND"
1062. .1856
/gene="spyM18_0956"
/notes="best blastp match gb|AAK33814.1| (AE006539)
histidine protein kinase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="histidine protein kinase"
/protein_id="AAL97597.1"
/db_xref="GI:19748122"
/translation="MPMTNQTLDLLDVYVYHAFRIAKALPNIPKTVLYLLEMLKE
RRELNLAFLAEHAAENRTIEDQTHCSLMLNQSLEDEQIANYILDLKVKVKGALIDFV
RSVPILYRLFLRLITSEIPNFKAYIFDKNQDQYDTHFQAMLESDEHVEFKAYLSQKQ
SRNVTTSLADMLTTLSPQEIKDVLRLHFEKAYRNPLAHLIKPFDDEELHRTTHF
SNOAFLENIITLATFSGVIVRREPFFVDDMNAIKKELSLWROSIV"
complement(1936. .2841)
/gene="cpsY"
/notes="spyM18_0957"
complement(1936. .2841)
/gene="cpsY"
/notes="best blastp match gb|AAK33815.1| (AE006539)
putative transcriptional regulator [Streptococcus pyogenes
M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL97598.1"
/db_xref="GI:19748123"
/translation="MRIQQLHYIKIVEGSMNEAAKOLFITOPSLNAVKDLEMENG
ITIFNRNKGITLTKDGVESLFYARQIIEOTSLLEDYRKNHNTGRELFVSSDHYAFV
VNAFVSLKRTMDTRELFTRETRWEIIDDVKNFSEIGVLFINDYNRDVLTKLFDD
NHUTASPLKAPQPHFVSKSNPLTKSLMDLDRDPFVLSYDQGHNSPFSEEMHS
OMPHKNSIVVSDRATFLNMLIGLDGTVASGILNSNLNGQIVAIPLDVPDEIDIVFI
KHEKANLSKMGERFIEVLEETFD"
3063. .3599
/gene="spyM18_0958"
/notes="spyM18_0958"
/notes="best blastp match gb|AAK33816.1| (AE006539)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97599.1"
/db_xref="GI:19748124"
/translation="MNVTEONSHQILKLLVSIHYLTFLRDELKLVERTPSILGCEP
PAHLVSELDIVAAIDLDMDQORLSETFWEEAFKLMNKZLDADYHNGKINGVDHLI
DLGOSKEVFOIIGDKRIRRVGLSDVFSLSKLSALSKLEAPIDPILYEHIQVRNLEE
EAFIKYQSPKVPTEGE"
3876. .4568
/gene="pyrF"
/notes="spyM18_0959"
3876. .4568
/notes="pyrF"
/notes="best blastp match gb|AAK33817.1| (AE006539)
putative orotidine-5'-decarboxylase PyrF [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative orotidine-5'-decarboxylase"
/protein_id="AAL97600.1"
/db_xref="GI:19748125"
/translation="MKEERPIIALDFSSFEETKAFDLDFPAEEKLYVKIGMELYAAG

```

PDIVRYIKSLGHNVFLDLKLDIPNTVRAAMAVLKELDDMATVHAAGGVEMLKAARE
GLQDPTLIAVLTOLTSSEDDMGDQNIQTSLESVHLVSKGAQAQLQDGVVCSAQEV
EAKAVPTGFTCLTGPRIKPGNSIGDQKRVMPNQARRIESDIIVVGRPIITQAKDPV
AAVOAIKAEWAG*
4626..5255
/gene="pyrE"
/note="spym18_0960"
4626..5255
/gene="pyrE"
/note="pyrE"
/note="best blastp match gb|AAK33818.1| (AE006539)
putative orotate phosphoribosyltransferase [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative orotate phosphoribosyltransferase"
/protein_id="AAL97601.1"
/db_xref="GI:19748126"
/translation="MTLASQIATQLLDIRAVLYLKPEDPFTWASGIKSPIYTONRVTL
YKTRDLIENGFEVETIRAHPEVEVIAGTATAGIPHGAIIDAKMTLPFAVIRSKPKDH
GACNOIEGRLKQKQWVILIEDLISVSGSVLDAAARAREGADVGVVAIFTYELPKAS
ONKEAGIKLITLSNTELIARVAKLOGIITNDGLHLLKFKEDQVNMQQ"
5452..6906
/gene="amiC"
/note="spym18_0961"
5452..6906
/gene="amiC"
/note="best blastp match gb|AAK33819.1| (AE006539)
putative amidase [Streptococcus pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="AAL97602.1"
/db_xref="GI:19748127"
/translation="MTYQDANMAIAVOTGQTPPLELVQAIYKAKKLNPTLNAITSE
RFEALAEAKQDFSGLPFAGVPIDFLDGLQELKHSSTSGSRLFKFYQNTKTDLFVK
RLGALGFILGRSNTPEFGKNIIDSLHGPVNLKPDNTRNAGSSGSGAAALVSSGIS
ALAPADSGGSIIRIPASFNGLIGLGRGMPVPGSYRSGWOGAFHATLSVRDTR
NLIYIQHOMESPFLAVLTKDSIQSQRPUIAFYORLSDGSPVSDTAKAKQA
VWLRGQHOLVEEFPVNMTEVIRHYIYMSVETAMFADIEDTLGRMKNDDMET
MTWAIYOSGKIDIPAWRYSOVLQKWDYTSATMASFHYETNLLLTFTTNPAPKHELVP
DKLMLANQAQELFSEEQNLVETMFGKSLAINPYTALPNLTGPAISLPTYETKEG
LPMGIOLIAAKGREDDLGLIAEQFEAAGLLKIPQ"
7024..7887
/gene="spym18_0962"
7024..7887
/gene="spym18_0962"
/note="best blastp match gb|AAK33820.1| (AE006539)
putative ABC transporter (binding protein) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (binding protein)"
/protein_id="AAL97603.1"
/db_xref="GI:19748128"
/translation="MKKSVTLLSIGLASLLAACAPHKSQKSSWNTIKKGVKLVKATP
GTQPTSPFNNDELGYEDVMWKEIGKRLNKKYFETGDFQAFVSDSGRVDISLN
NFDITPARKKYNINISPYIGYGMIVRADGSSNAKKDLSDMKGKAAGASGETYMK
VAQQAENVYDNYDVLNDVANGRTDFIPNDYPAQKLFVDYMLSQNPRLNVKMS
DVQNPTEQIGVNNKDDSLKKIDAVIKDMKKDGLKKIKSETIYVAGQDLTKPFGKDK
KIPVDTKDVN"
7916..8566
/gene="spym18_0963"
7916..8566
/gene="spym18_0963"
/note="best blastp match gb|AAK33821.1| (AE006539)
putative ABC transporter (permease) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (permease)"
/protein_id="AAL97604.1"
/db_xref="GI:19748129"
/translation="MINIPLMKDSLGLVLSGLUPYTLIGISULLSFLTGLVGLLLALLGR

SRPOLLHYLVRTYISIMRGVPMIVLVFLVLYFGLPYGCLPALLCAYLGFWSVSAAYI
SEVFRSIIAIDKQWEAAKALGLPILMVAIYLPQAFRIAVPPVGNVLIIDWVSSS
LAAMITVPDIFQNAKIIGRENDYMSYILVAFIYLIAPFLERYOEFLKALV*
8699..9352
/gene="ung"
/note="spym18_0964"
8699..9352
/gene="ung"
/note="best blastp match gb|AAK33822.1| (AE006539)
Query Match 42.9%; Score 275.6; DB 1; Length 11573;
Best Local Similarity 64.3%; Pred. No. 7.8e-59;
Matches 413; Conservative 0; Mismatches 229; Indels 0; Gaps 0;
QY 1 ATGATTACAATAGTTTATTAATCCTAGCTATCTGCTGGTTCGATTCATCCTGGTCTC 60
DB 11478 ATGAATTTACTACTTTTATTTATACCAATTCCTATTTACTTGGTCTTATTCACATGGA 11419
QY 61 TGGNTTGGCAAGATTTCTTCAANTCAATCTACCGCAGCATGTTCTGTTAACAACATGGA 120
DB 11418 TGGATTGGCAGAGTACTTTTACCACATCACTTACGAGAGCATGATCAGGAAATACCTGGA 11359
QY 121 ACGACCAACACCTTCGCGATTTTAGTGAAGAAAGCTGGTATGCAACCTTTTGTGATTGAC 180
DB 11358 ACCCAATATCTTTTCGATTTTAGGTGTCAGGCGAGCAACACTTCTAGCTATTGAT 11299
QY 181 TTTTCAAAGGAACCTTAGCAAGCTGCTTCCGATTTATTTTTCATCTACAAGCGTTTCT 240
DB 11298 ATGTTTAAAGGGACACTTTTCAATATGTTTACCAATATTTTGTGATGACTTCAATTTCA 11239
QY 241 CCTCTCATCTTTGACATTTTGGCTGTTATCGGCCATACCTTCCTCTATCTTTTGACAGATT 300
DB 11238 TCCATTGCTATCGGCTTTTTCGAGTTTTCAGGCATACCTTTCTCTATTTTGGCACTTT 11179
QY 301 AAAGTGGTGAAGCTGTCGCAACAGTGTGGAGTGAATTTTCGGATTTTCGCCCTATCTTC 360
DB 11178 AAAGTGGTGAAGCTGTCGCAACAGTGTGGAGTGAATTTTCGGATTTTCGCCCTATCTTC 11119
QY 361 TGTCTCTACCTTCGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
DB 11118 CTCTTTTTCATCATCTATCTTTGTTTATTTAGTCTCTCTATTTATTTAGTATGATATCTTA 11059
QY 421 TCTAGTGTACACATCGATTGCGGCTGTTATCGGGTTCCTGCTCTTTTCCACTTTTGGT 480
DB 11058 GCTAGTGTGCTTTTCAGCATCTGTTGGTGTGTTATCTGTTTAAACATTTTCTGCGCATTCAT 10999
QY 481 TTTTCTCTAGTAACTATGATCTCTCTCTCATCGCTATTTATCTTAGCAGCTTGTAGTTTG 540
DB 10998 TTTCTTTTACCAATATGACTACTTTTAACTTTTATGATATGATTTGCGTTCATT 10939
QY 541 ATTATCATTCGTATAGGACAAATATAGCTCGTATCAAAAATGAAATGAAATTTGGTC 600
DB 10938 ATTATTAAGACACAAAGATAACATTAGTCGTATCAACATCATCTACTGAAATCTAATA 10879
QY 601 CCTTGGGATTAACCTTAACCCATCAAGATCCTTAAATAATAA 642
DB 10878 CCTTGGGGCTAAATTTAAGCAACAAGTGCCTTAAATAATAA 10837
RESULT 9
AE014149/c
LOCUS
DEFINITION
Strepptococcus pyogenes MGAS315, section 14 of 37 of the complete
genome.
ACCESSION
AE014149
VERSION
AE014149.1
KEYWORDS
GI:21904329
SOURCE
Strepptococcus pyogenes MGAS315.
ORGANISM
Strepptococcus pyogenes MGAS315
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 53354)
AUTHORS
Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S.,

TITLE	Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.	CDS	/gene="clpX" /note="SpyM3_0604" 1883..3112 /gene="clpX" /note="best blastp hit: gb AAK33805.1 (AE006538) putative ATP-dependent Clp protease subunit X [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative ATP-dependent Clp protease subunit X" /protein_id="AAW79211.1" /db_xref="GI:21904331"
	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002) 2 (bases 1 to 53354) Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.		
TITLE	Direct Submission Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis, Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St., Hamilton, MT 59840, USA Location/Qualifiers 1..53354 /organism="Streptococcus pyogenes MGAS315" /strain="MGAS315" /serotype="M3" /db_xref="taxon:198466" /note="group: A" 123..986 /gene="thyA" /note="SpyM3_0601" 123..986 /gene="thyA" /note="best blastp hit: gb AAK33802.1 (AE006538) putative thymidylate synthase [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative thymidylate synthase" /protein_id="AAW79208.1" /db_xref="GI:21904330"	gene	/translation="MAGSRNDIKVYSCFGKSODDVKKIAGNNVVICNECVASLSE LIKEELAEVLADLTPVKPKELDVNOYVVGODPAKRALSVAVNVHVKVSEFESR DDDDVLDQSRNLMIGPTSGKTFEATLAKSLNVPFAIADATSLTEAGYVGEDVENI LKLQAAADYNVERAERGIIYDEIDKIAKGENSVITRDVSGEGVQVQGLKIIETIV ASVPPGGGRKHNMEOIMIDTKNLPFVIGAGFDGIEEIVQRLGEKVIQFGQNSRKID DNASYMOEIISEIDQIFGLIPEFIPYVALEQLNTSLDILQILTEPRNALVKYQQA LLSYDGVLEAFDKEALEAIANKAIERKTGARGLSRIETETMLDIMEIPSQEDVTAKR ITKAAVEGSKPKVLETA" 3122..3721 /gene="SpyM3_0605" 3122..3721 /gene="SpyM3_0605" /note="best blastp hit: gb AAK33806.1 (AE006538) conserved hypothetical GTP-binding protein [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative GTP-binding protein" /protein_id="AAW79212.1" /db_xref="GI:21904334"
	translational="MAEQVLTNHNASILLSSANKSHYPQDDLPETALAGRSNVGKSS FINTILGRNKLARTSKPGKQTLLNFNIDDKLRFVDYGYGAKVSKSERAKWGKMI EYLTLGRNLRVAVSLVDRHAPSKEDIQYDFLKYIDPVIIVVATKADKIPHGKWNK HESVVKALNFDKSDTFIVFSSVERIGIDSDWAILEQV" 3869..4615 /gene="SpyM3_0606" 3869..4615 /gene="SpyM3_0606" /note="best blastp hit: gb AAK33807.1 (AE006538) hypothetical protein [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAW79213.1" /db_xref="GI:21904335"		
TITLE	SKYITGAFAYDLAKGEFPITLRPIKSAIKLEFWIYODQNSLDVLEAKYNVHW NEWEDOTRIGORYGAVYKXKDIISKILQLAENPWRNRNVLNDYEFETKGLL PCAFQIFDVRVGEDYLDASLTSRNDILVAHHINAMQYVALQMMIAKHPGWKIG FFYFVNNLHYDQFOAQLLRQVPASQPKLVNLVNPDRNFFDIKPDQDFELONYDP VKPOLHFDLAI" 1066..1563 /gene="dyr" /note="SpyM3_0602" 1066..1563 /gene="dyr" /note="best blastp hit: gb AAK33803.1 (AE006538) putative dihydrofolate reductase [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative dihydrofolate reductase" /protein_id="AAW79209.1" /db_xref="GI:21904331"	gene	/translation="MTQVAQGRSGKSLCHOCNALFVFIILILLAVLFGVYRNS IEGIWPTTSIDOKLGDGDFAKRLTGLHQSPLIDDSLTSSOMILTYKNNVNDLSFSVOV ERDIFVKRLAANYHONELKTLKHNHLVYVGLDLSKQRIIENSHPASHELEMLDQAFE KLASQIGGRYNORTGLHSVAVLKVNRILHTIDIKEEVAAGHTSFSGLLTPNGIFYD YTRFGKLELLAGDEKTIFFKALKKSVKCI" complement(4668..6767) /gene="clpX" /note="SpyM3_0607" complement(4668..6767) /gene="clpX" /note="best blastp hit: gb AAK33808.1 (AE006538) putative ATP-dependent Clp proteinase (ATP-binding subunit) [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative ATP-dependent Clp proteinase (ATP-binding subunit)" /protein_id="AAW79214.1" /db_xref="GI:21904336"
	MGRVTFEGMCKRLPORQTLVMTNRDRYQVDEVLMTSIEKVLWEVHAQDKTLYIIGG NKVLEAFNGYFDRIKTVIHHRFGDGYRPNLDFSHFTQESQFYARDAKNPYDFVT ALXHK" 1583..1753 /gene="SpyM3_0603" 1583..1753 /gene="SpyM3_0603" /note="best non-GAS blastp hit: gb AAK75657.1 (AE007451) hypothetical protein [Streptococcus pneumoniae TIGR4], and gb AA00232.1 (AE008512) hypothetical protein [Streptococcus pneumoniae R6]" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAW79210.1" /db_xref="GI:21904332"		
TITLE	VNSIHAFTGLF" 1883..3112	gene	/translation="MQRSEFGVTFALVCIVLCVCAIPAFKKNRYGLASFLFLNAFTNL 1883..3112


```

/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase alpha chain
(EC 1.17.4.1)"
/protein_id="AAK05073.1"
/db_xref="GI:12723918"
/translation="MSLQKQDVYFYFKLNNEIPVNGSIPLAKDKEAITAFFKENVE
PNRYEADTYIDKLNWLKVEYLEAEFLKYKPEFTNLRKELSKFFHFDSFMAAYKF
NQYAKNNGDSLYLEDPEDRVLMNALYMDGDEKLANNLAFAMINRYPATPTEFL
AGRRGEFYSCFLQLQUTDDMSIGRTINSALQSLKNGGVLNLTNLRAGAPITGI
DGAAGVVPYKMLFEDAFSYANGLOGRQAGVYLSLFHPDIMEFTSKKENADEKIR
KTKSLGATVPYDFKVELVKGEMTYLFEYFVEKYKGPFAEVDITAELYDKMANPEI
RKTISNARELEQELSKLOESGYPYIVNVTANOANPDVGISMSNLCSEILOVQTPS
VLNDQTYKVVGDVACNLGTSNTVNNMTSSNEEFTSIEAMVRLTFISETSLDTPV
TVRGKNDMAHGLGAMGLHSFLAQNHQHYDSKEAVEFTSYFMLNLYTLVASNKLA
VEKGSKNFEKRAYADGSYDFKYLTHDYFAVSSKVQELFGIAVPTIEDTALRDT
VETSGLYNSYRMVAVPNGSISYINDCSSIHPINRIEERQEKVKGIYYPAAGLATE
TPIYKAYOTDMRAVIDYVAAATEHVDQGLSLTLPMSRSTLPEGLYEMKVANNKMTTR
DLSLRNAYKKGITYYIRTFTDQEEVGANQCESCVI"
complement(3713..4135)
/gene="ndrI"
/notes="L4702"
complement(3713..4135)
/gene="ndrI"
/notes="EVIDENCE BY HOMOLOGOY BIO06.05 ENERGY METABOLISM
Electron transport. BELONGS TO THE NRDI FAMILY"
/codon_start=1
/transl_table=11
/product="ribonucleotide reductase"
/protein_id="AAK05074.1"
/db_xref="GI:12723919"
/translation="MKLAFESVGTQTRFVSKTNLNPVELSPDDIEMNEPFLITPS
YAESPVKSIDVPDFMAYNONYKHCLGIGTGGRNFAGIYITAKELSAKIQ
IPLYDFENGTPADVAAEKELATQLDKGAKVTFKNPL"
complement(4137..4313)
/gene="ndrH"
/notes="L5126"
complement(4137..4313)
/gene="ndrH"
/notes="EVIDENCE BY HOMOLOGOY BIO06.05 ENERGY METABOLISM
Electron transport. ELECTRON TRANSPORT SYSTEM FOR
RIBONUCLEOTIDE REDUCTASE SYSTEM NRDEF"
/codon_start=1
/transl_table=11
/product="glutaredoxin-like protein NrdH"
/protein_id="AAK05075.1"
/db_xref="GI:12723920"
/translation="MVKKWLSEHEIAFDEINIDEQPEFVEKVIEMGFRAAPVITKDDF
AFSGFRPSELAKLA"
complement(4529..5170)
/gene="ykcC"
/notes="L5517"
complement(4529..5170)
/gene="ykcC"
/notes="EVIDENCE BY HOMOLOGOY BIO15.02 HYPOTHETICAL.
Conserved. integral membrane protein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05076.1"
/db_xref="GI:12723921"
/translation="MLTIILLIASYLLGAIPFLGLWTGKTIFFKKNLHDYSGNVTGTTN
TFRILGVKAGISVAFDLKGLATLPLFFHNGVSPLIPLGLAVIGHGTFISYDFRKF
GGRAVATVSGVLSFPLFIYLLVWFIVWLFSMISLSVIGAVFALLGIILFIPSI
GFIITSDYDLFSIIIFVLAIIIIILRHRTNLKRKNHCESLVPEGLNLSKOKER"
5445..7379
/gene="parE"
/notes="L0290"
5445..7379
/gene="parE"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit B (EC 5.99.1.1-)"

```

```

/protein_id="AAK05077.1"
/db_xref="GI:12723922"
/translation="MWIDINNYDDSAIQVLEGLDANVRKRPCTWYIGSTDGTGLHLHWE
IVDNAVEALSGFNGRISVITINADGSLSVSEDEGRMPVGHKAMGIPTEVFITLHAG
KFGQGGYKTSGLGVGSSVNALSTWLEVEITRDGTVYRORFENGHPATTLVKIG
KAPKSTGTVKTEPMDSSIFSTTEFYNIISERLNSFAFLKEVITISLEDKRDEKEE
FHVENGQDPSVYLNDDKSLTPVLYPEGQDGFVEVALQYNDGYSENTLSFVNNVR
TKDGGTHEAGLTAITKAMNEYARKVNLKKEKDNLEGSYREGLTVVLSVLPEEHL
QFSGQTKDLGSLARPVVDLSYSEKLSFYLMENGELAQNLIIRKAIKARARAAKKA
REDNRGKSKKDGSLLSGLTTPAOTKNPNKNIILYVEGDSAGSAGAKORHAKFOAIL
PLRGKVINTEKAKHQDILKNEETIMTYITIGAGVADFTLADRNKYDKIIIMTDADTDG
AHQITLLTTFYRMPKPLVEAGHVIYALPLYKMSGCKGKKIEIYAWTQGLAEALRK
QFQKNANLQRYKGLGEMNADQLWETMDPETRTLIOVTTIDDAQAQAEKRVSLMGDKVE
PRKRTIENNVKFTMEETVE"
7584..7790
/gene="ykaE"
/notes="L8551"
7584..7790
/gene="ykaE"
/notes="EVIDENCE BY GENEMARK BIO16.01 UNKNOWN"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05078.1"
/db_xref="GI:12723923"
/translation="MOKSALIIDILOAFIKKLAKDAVPYGDWIKIENNRALISAFDOK
NQPYIIVTEPKYFGNQLKALFLG"
7787..8122
/gene="ykaF"
/notes="L8730"
7787..8122
/gene="ykaF"
/notes="EVIDENCE BY HOMOLOGOY BIO15.01 HYPOTHETICAL"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05079.1"
/db_xref="GI:12723924"
/translation="MILQDELDKNRLHQLVKYGPSAFKOSDYGLADELKNNSVKIY
VSGVSHSNGVLTKARDGADLGFEAMIEDACDRKLKNFQKTKTDEIPKFGKLIKTE
IIEELKNV"
8115..8948
/gene="dnaQ"
/notes="L9094"
8115..8948
/gene="dnaQ"
/notes="EVIDENCE BY HOMOLOGOY BIO10.02 DNA replication,
restriction, modification, recombination, and repair. DNA
polymerase III responsible for most of the replication
synthesis in bacteria. Exhibits also 3' to 5' exonuclease
activity."
/codon_start=1
/transl_table=11
/product="DNA polymerase III, epsilon chain"
/protein_id="AAK05080.1"
/db_xref="GI:12723925"
/translation="MSEKYVVFDFETTFGSAQKNEIIQIGAVKYDENNQDQARFNOLY
KNORSYSTEITQIGTKPADLNOPTIEILPEFLDFIKGHLIVAHNAPFDSIFYQ
AIIDCAITDVEAFRVDTLAESKKLLTMSYALENKKLLEMDLRSHDNLCLITQA
KLQYIQLKIEVELSEVSEVTDDELGDLDFGSDSDEITDSLRKLNLPITKALVYQO
NLKRWREVLGISIEREYVSGSSNLRLHNSKVKIHSDFLEKMQKANFVTIMEEE
K"
8945..9394
/gene="ykbA"
/notes="L9876"
8945..9394
/gene="ykbA"
/notes="EVIDENCE BY HOMOLOGOY BIO15.01 HYPOTHETICAL.General"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05081.1"
/db_xref="GI:12723926"

```

/translation="MTAIWKEKSFNQLTLDYVOLLRLTAVFVVEONCPVOELDKGD
QVAKHLWENHGMALCRLPKGISFAESIGRVLNPEFRQNLGRGLIALEV

Query Match 38.4%; Score 246.6; DB 1; Length 13211;
Best Local Similarity 61.7%; Pred. No. 1.5e-51;
Matches 393; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 2 TGATTTACAAATAGTTTATTAATCCTAGCTATCTGCTGGTTCGATTCATCTCGTCTCT 61
DB 5166 TTATATTTATTTCTTCTCTCAATGCCAGCTATCTTTTGGGGCAATTCATTTGGCTTAT 5107

QY 62 GGATTTGGACAACTATCTTCAAACTACGCGACATGGTCTGTAACACTGCAA 121
DB 5106 GGATTTGGCAAGATTTTTCAAAAAATCTCCATGACTATGGCTCAGAAATACTGGA 5047

QY 122 GCACCAACACCTTCGCCATTTAGTAAGAAGCTGGTATGCCAACCTTTTGTGATGACT 181
DB 5046 CAACCAATATCTTTAGAAATCTAGCTGTAAAGGCAAGGAATTCAGTTTTCGCTTTGACT 4987

QY 182 TTTTCAAGGAACCTAGCAAGCTGCTCCGATTTATTTCACTACAGCGCTTTCTC 241
DB 4986 TACTAAAGGGGACACTTCTACTCTTTTACCCCTTTCTTCCATATCAATGGTGTTCAC 4927

QY 242 CTCTCATCTTTGGACTTTTGGCTGTATCGGCATACCTTCCCTATCTTTGGAGTATTA 301
DB 4926 CTCTAATTTTGGCTTCTTCTGCTGCTATGACATCTTTTCAATATTCACCGCTTTA 4867

QY 302 AAGTGGTAAGGCTTCGCAACCACTGCTGGAGTATTTTCGGATTTTCGGCTATCTTCT 361
DB 4866 AAGTGGTAAGGCTTCGCAACCTGCTGGTGTGATTTTGGATTTAGTCCGCTTCT 4807

QY 362 GTCCTACTCTGGATATCTTCTTTGGAGCTCTATCTTTGGGAGTATGATTTTCACTGT 421
DB 4806 TAATTTACTCTTCTTCTCTTATCATGTCCTATGCTTTTATGACATGATTTAGTCTCT 4747

QY 422 CTAGTGTACAGCATCGATTCGGCTGTATTCGGGGTTCTGCTCTTCTCCACTTTTGGTT 481
DB 4746 CTAGTGTAGTCTGTATTTGCTTTATTTGGGAATCTTATTTTCCCTAGCATCGAT 4687

QY 482 TTATCCTGAGTAATCATGACTCTCTCTCATCGCTATTTATCTTTAGCACTTCTAGTTTGA 541
DB 4686 TTATTTCTTACTTCTTACGACCTTTTATCTCAATTTATTTATTTTGTGCAATTTATTA 4627

QY 542 TTATCATCTGTCATAGGACATATAGCTCTCATCAAAATTAATAAATAAATTTGGTCC 601
DB 4626 TTATTTTACGACATAGACAAATCTCAAAAGAACTCAAAATCATTTGCGAATCTTGTCC 4567

QY 602 CTTGGGGATTGAACCTAACCCATCAAGATCCTAAAAA 638
DB 4566 CATTTGGTTTAACTTTATCAAAACAAGAAATA 4530

RESULT 11
SGO236899 771 bp DNA linear BCT 05-APR-1999
LOCUS Streptococcus gordonii parE (partial) gene and ORF2 (partial).
DEFINITION
ACCESSION AJ236899
VERSION AJ236899.1 GI:4584071
KEYWORDS hypothetical protein; ORF2; parE gene; topoisomerase IV.
SOURCE Streptococcus gordonii.
ORGANISM Streptococcus gordonii.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 771)
AUTHORS Vriesema, A.J., Dankert, J. and Zaai, S.A.
TITLE Isolation and characterization of promoter regions from Streptococcus gordonii CH1
JOURNAL unpublished
REFERENCE 2 (bases 1 to 771)
AUTHORS Vriesema, A.J.M.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1999) Vriesema A.J.M., Medical Microbiology, Academic Medical Center, University of Amsterdam, Room L1-160,

Meibergdreef 15, 1105 AZ Amsterdam, NETHERLANDS
FEATURES
source location/Qualifiers
1. 771
/organism="Streptococcus gordonii"
/strain="CH1"
/db_xref="taxon:1302"
complement(1..176)
/gene="parE"
complement(<1..104)
/gene="parE"
/codon_start=1
/transl_table=1
/product="topoisomerase IV"
/protein_id="CAB40548.1"
/db_xref="GI:4584072"
/translation="MAKKEININNYNDATQVLEGLDAVRKRPQWYIGS"
complement(111..116)
/gene="parE"
complement(150..155)
/gene="parE"
complement(171..176)
/gene="parE"
202..207
/note="ORF2"
224..229
/note="ORF2"
267..>771
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAB40549.1"
/db_xref="GI:4584073"
/translation="MINTILGLILAYLLGSIPITGLWIGQIFPKKRLREYSGNTGTTN
TFRILAKTAGTDFATDFLKGTLATLLPLFLHNGISPMIFGLIYLVGHFPPIFAEFK
GGRVATSGVWAGFSPFLSYLIIIFIVTLXLSMISLASIVVAGFAIISLVILFPL
GIILPSVD"
BASE COUNT 215 a 147 c 121 g 287 t 1 others
ORIGIN
Query Match 38.1%; Score 244.6; DB 1; Length 771;
Best Local Similarity 68.0%; Pred. No. 5.7e-51;
Matches 340; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 1 ATGATTTCAATAGTTTATTAATCCTAGCTTATCTGCTGGTTCGATTTCCATCTGCTC 60
DB 270 ATAAATACAAATTTGGGCTAATTTTGGCATATTTACTGGTTCATTTCCAACTGGTTA 329

QY 61 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCGAGCATGCTTCTGGTAACTGGA 120
DB 330 TGGATTGGACAGATTTCTTTTAAAGAAGATTTACGAGAATACGGTAGTGGAAATACGGGC 389

QY 121 ACACCAACACTTCCGCAATTTAGTAAGAAGACTGCTATGGCAACCTTTGTGATTGAC 180
DB 390 ACAACCAACACTTCCGCTATTTTAGGAAGACACTGGAAGTATGATTTTGGCAATCGAC 449

QY 181 TTTTCAAGGAACCTTAGCAACCTGCTTCCGATTTATTTTTCATCTACAAGGCTTTCT 240
DB 450 TTTTCAAGGAACCTTTGCAACCTTACTTCCACTTTTCTCCATATCAATGATATCA 509

QY 241 CTTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGGAGGATTT 300
DB 510 CCAATGATTTTGGTCTTAATAGCTGTTCTTCCACACCTTTCCCAATTTTTCGTAATTT 569

QY 301 AAAGGTGTAAGGCTGCGCAACCACTGCTGGAGTATTTTTCGATTTTGGCCCTATCTTC 360
DB 570 AAGGAGCAAGGCGGTTGCTACAAGTGCAGGCTGTTGTCNTGGATTTTCTCCCTCTTC 629

QY 361 TGTCTCTACCTGGATATCTTCTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCTACTG 420
DB 630 TTTTCCATCTGATTTATTTATTTATGTAACGCTTTACTTTGGGAAGTATGATTTCTTTG 689

QY 421 TCTAGTGTACAGCATCGATTTGGGCTGTTATCGGGGTTCTGCTCTTCTCCACTTTTGGT 480

BASE COUNT 57412 a 39214 c 31781 g 66862 t
ORIGIN

Query Match 19.3%; Score 124; DB 6; Length 195269;
Best Local Similarity 56.1%; Pred. No. 8e-21;
Matches 257; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

QY 1 ATGATTACAATAGTTTATTAATCCTAGCCCTATCTGCTGGGTTGCAATTCATCTGGTCTC 60
DB 117156 ATTAATTTAAATTTTGGCTTTCTTGTAGCTTATGTAATCGGTCAATACCTTCTGGCTTA 117097

QY 61 TGGATTGACAAAGTATTTCTTTCAATCAATCTACCGAGACATGGTTCTGGTAACACTGGA 120
DB 117096 TGGATCGGTAAATTTTACAAAAAAGATATTCGTGATTTTGGTAGCGGAATTTAGGA 117037

QY 121 ACGACCAACACCTCCCGCATTTTAGGTAAGAAAGCTGCTATGGCAACCTTTGTGATTGAC 180
DB 117036 GCTACAAACTCTTCCCGTGTTTTAGGGGTAAGAGCGGCAGTATTTGTACAGTGATGAT 116977

QY 181 TTTTCAAGAACCCCTAGCAACGCTGCTTCCGATTTATTTTCACTACAAGGCGTTTCT 240
DB 116976 ATTTTAAAGGTACTGTCGCAACATTACTTCCATTCTTTTCCAACTAAATGTTAATCAT 116917

QY 241 CCTCTCATCTT-----TGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTGCA 294
DB 116916 CATTTCTGGTTACTTACGGGAGCATTTTGGCATTTATCGGACATAGTTTCCCGCTTTTCGA 116857

QY 295 GGATTTAAAGGTGGTAAGGCTGTCGCAACAGTCTGCTGAGTGATTTTCGGGATTTGCGCCT 354
DB 116856 GGATTCAGAGTGGGAAGAGAGTGTCTACTTCTGCTGGTGTATCCTTGTCTTATGACCA 116797

QY 355 ATCTTCTGCTCTACCTTTGCGATTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATT 414
DB 116796 CTTTGTGTTGTAGCAGCACTAGTCGTTTCTTATTAAACATTAATAATCAGCAATATCTG 116737

QY 415 TCACCTGCTAGTGTACACAGCATCGATTGCGGCTGTTAT 452
DB 116736 TCGCTTAGTTCAATGATTGGAGCACTTGCAGCATTAAT 116699

Search completed: December 25, 2002, 10:43:28
Job time : 11907.8 secs

RESULT 15
AX417044/c 349980 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 4035 from Patent WO0228891.
ACCESSION AX417044
VERSION AX417044.1 GI:21449654
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4035 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1. 349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="seq 2058, original length: 3,011,208 replaced
by-seq 2058: 0,000,001 to 0,349,980-seq 4032: 0,300,001 to
0,649,980-seq 4033: 0,600,001 to 0,949,980-seq 4034:
0,900,001 to 1,249,980-seq 4035: 1,200,001 to
1,549,980-seq 4036: 1,500,001 to 1,849,980-seq 4037:
1,800,001 to 2,149,980-seq 4038: 2,100,001 to
2,449,980-seq 4039: 2,400,001 to 2,749,980-seq 4040:
2,700,001 to 3,049,980-seq 4041: 3,000,001 to 3,011,208"
BASE COUNT 116021 a 60232 c 66706 g 107021 t
ORIGIN

Query Match 19.3%; Score 124; DB 6; Length 349980;
Best Local Similarity 56.1%; Pred. No. 7.7e-21;
Matches 257; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 150.797 Seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	642	100.0	642	21	AAA09181
2	642	100.0	6171	19	AAV52170
3	640.4	99.8	642	23	AAV55594
4	640.4	99.8	642	23	AAV55801
5	327.8	51.1	344	20	AAZ20352
6	284	44.2	636	24	ABN67449
7	276.2	43.0	621	24	ABN70980
8	272.6	42.5	720	24	ABN67450
9	193.2	30.1	6691	20	AAI13025

10	124	19.3	495269	24	ABQ67195
11	102.4	16.0	615	24	ABN92939
12	102.4	16.0	818	18	AAV75059
13	102.4	16.0	3014	22	AAH54476
14	102.4	16.0	4105	22	AAH54115
15	100	15.6	582	20	AAZ20371
16	100	15.6	582	21	AAA09182
17	99.2	15.5	606	23	AAV51668
18	99.2	15.5	606	23	AAV54402
19	96	15.0	594	22	AAH53731
20	67.4	10.5	600	23	AAV53263
21	67.4	10.5	1830121	17	AAV74063
22	51	9.5	1033	20	AAV38290
23	59	9.2	570	23	AAV54065
24	57.8	9.0	612	23	AAV56022
25	56.6	8.8	189	23	AAV50887
26	53.6	8.3	618	23	AAV52569
27	49.4	7.7	1194	23	AAV86672
28	49.4	7.7	1194	23	AAV86672
29	40.4	6.3	5059	20	AAV84332
30	39	6.1	7758	24	ABL33103
31	38.6	6.0	10369	22	AAV46304
32	38.6	6.0	10369	24	ABL32393
33	38.4	6.0	618	23	AAV53979
34	38.4	6.0	16602	24	ABN80069
35	38.4	6.0	16602	24	ABL32727
36	38.2	6.0	13606	22	AAV54557
37	38.2	6.0	13606	22	AAV46561
38	38.2	6.0	13606	24	ABL33810
39	38.2	6.0	13606	24	ABK28313
40	38	5.9	781	24	ABQ14544
41	38	5.9	781	24	ABQ14545
42	37.6	5.9	522	21	AAV53681
43	37.4	5.8	32186	22	AAV34422
44	37.4	5.8	102634	21	AAV81464
45	37.4	5.8	349980	21	AAV21609

ALIGNMENTS

RESULT 1
AAA09181
ID AAA09181 standard; DNA; 642 BP.
XX
AC AAA09181;
XX
XX
DT 10-AUG-2000 (first entry)
XX
DE S. pneumoniae S-ynes coding sequence.
XX
KW S-ynes; survival; antibacterial; inhibitor; ds.
XX Streptococcus pneumoniae.
OS
PN WO200020527-A1.
XX
PD 13-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22665.
XX
PR 30-SEP-1998; 98US-0163445.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Fritz C, Youngman P, Guzman L;
XX
DR WPI: 2000-303799/26.
DR P-PSDB: AAV92246.
XX
PT Methods for identifying an antibacterial agent for treating
Streptococcus pneumoniae infections comprises detecting an interaction
between a ynes polypeptide and a test compound

Listeria innocua c
Staphylococcus epi
Staphylococcus aur
S. epidermidis gen
S. epidermidis gen
B. subtilis B-ynes
B. subtilis B-ynes
Staphylococcus aur
Staphylococcus aur
S. epidermidis ope
Haemophilus influe
Haemophilus influe
L. lactis DNA cont
Pseudomonas aerugi
Salmonella typhi D
Staphylococcus aur
E. coli DNA for ce
DNA encoding novel
DNA encoding novel
Stealth virus nucl
Human immune syste
Tumour suppressor
Human immune syste
Klebsiella pneumon
Human chemically m
Human immune syste
Chemically pretrea
Tumour suppressor
Human immune syste
DNA transcriptions
Oligonucleotide fo
Oligonucleotide fo
Heisseria meningit
Human DNA for a no
N. meningitidis pa
Neisseria meningit

XX Disclosure: Fig 1; 65pp; English.

XX Streptococcus pneumoniae yneS gene encodes a polypeptide (S-yneS) essential for survival for a wide range of bacteria. Identifying an CC antibacterial agent comprises contacting a yneS polypeptide (S-yneS) with CC a test compound and detecting an interaction of the test compound with CC the S-yneS polypeptide which indicates that the compound is an CC antibacterial agent. Alternatively, detecting a decrease in function of CC the polypeptide contacted with the test compound and determining whether CC the compound inhibits growth of bacteria, relative to the growth of CC bacteria cultured in the absence of a test compound where inhibition of CC growth indicates the compound is an antibacterial agent. Inhibitors of CC S-yneS function are useful for treating a Streptococcus pneumoniae CC infection in mammals.

XX SQ Sequence 642 BP; 151 A; 143 C; 122 G; 226 T; 0 other;

Query Match 100.0%; Score 642; DB 21; Length 642;
 Best Local Similarity 100.0%; Pred. No. 3.8e-178; Indels 0; Gaps 0;
 Matches 642; Conservative 0; Mismatches 0;

QY 1 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 60
 DB 1 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 60

QY 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGACATGTTCTGGTAACACTGGA 120
 DB 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGACATGTTCTGGTAACACTGGA 120

QY 121 AGCACCACACCTTCGCGATTTTGTAGTAAGAAAGCTGGTATGCGAACCTTCTGATTGAC 180
 DB 121 AGCACCACACCTTCGCGATTTTGTAGTAAGAAAGCTGGTATGCGAACCTTCTGATTGAC 180

QY 181 TTTTCAAGAGAACCTTAGCAAGCTGCTCCGATTTATTTTCATCAAGCGGTTTCT 240
 DB 181 TTTTCAAGAGAACCTTAGCAAGCTGCTCCGATTTATTTTCATCAAGCGGTTTCT 240

QY 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300
 DB 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300

QY 301 AAAGTGGTAAGCTGTCGCAACCACTGCTGGAGTATTTTCGGATTTTCGCGCTATCTTC 360
 DB 301 AAAGTGGTAAGCTGTCGCAACCACTGCTGGAGTATTTTCGGATTTTCGCGCTATCTTC 360

QY 361 TGTCTCTACCTTGGCGATTAATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTTCAC 420
 DB 361 TGTCTCTACCTTGGCGATTAATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTTCAC 420

QY 421 TCTAGTGTACAGCATCGATTGGCGCTGTTATCGGGGTTCTGCTTTCACCTTTTGGT 480
 DB 421 TCTAGTGTACAGCATCGATTGGCGCTGTTATCGGGGTTCTGCTTTCACCTTTTGGT 480

QY 481 TTTATCTCAGTAACATATGACTCTCTCTTCATCGCTATTTATCTTAGCACTTCTAGTTTG 540
 DB 481 TTTATCTCAGTAACATATGACTCTCTCTTCATCGCTATTTATCTTAGCACTTCTAGTTTG 540

QY 541 ATTATCATCTTGCATTAAGACAAATATAGTCTGTATCAAAAAATAAACTGAAAAATTTGGTC 600
 DB 541 ATTATCATCTTGCATTAAGACAAATATAGTCTGTATCAAAAAATAAACTGAAAAATTTGGTC 600

QY 601 CCTTGGGGATTGAACCTTAACCCATCAAGATCCTTAATAATAA 642
 DB 601 CCTTGGGGATTGAACCTTAACCCATCAAGATCCTTAATAATAA 642

RESULT 2
 AAV52170
 ID AAV52170 standard; DNA; 6171 BP.
 XX
 AC
 AAV52170;
 XX

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:37.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 PN WO9818931-A2.
 XX 07-MAY-1998.
 PD 30-OCT-1997; 97WO-US19588.
 PF 31-OCT-1996; 96US-0029960.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX WPI; 1998-272225/24.

DR Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 XX pneumoniae

PS Claim 1; Page 371-374; 1409pp; English.

XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridize to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX SQ Sequence 6171 BP; 1645 A; 1424 C; 1188 G; 1914 T; 0 other;

Query Match 100.0%; Score 642; DB 19; Length 6171;
 Best Local Similarity 100.0%; Pred. No. 9.8e-178;
 Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 60
 DB 5149 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 5208

QY 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGACATGTTCTGGTAACACTGGA 120
 DB 5209 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGACATGTTCTGGTAACACTGGA 5268

QY 121 AGCACCACACCTTCCGATTTTGTAGTAAGAAAGCTGGTATGCGAACCTTCTGATTGAC 180
 DB 5269 AGCACCACACCTTCCGATTTTGTAGTAAGAAAGCTGGTATGCGAACCTTCTGATTGAC 5328

QY 181 TTTTCAAGAGAACCTTAGCAAGCTGCTCCGATTTATTTTCATCTACAGCGGTTTCT 240

Db 5329 TTTTCAAGAACCCCTAGCAAGCTGCTCCGATTATTTTTCATCTACAAGCGTTTCT 5388
Qy 241 CCTCTCATCTTTGGACTTTGGCTGTTATCGGCCATACCTCCCTATCTTTGCGAGGATTT 300
Db 5389 COTCTCATCTTTGGACTTTGGCTGTTATCGGCCATACCTCCCTATCTTTGCGAGGATTT 5448
Qy 301 AAAGTGCTAAGGCTGTCGCAACCAAGTCTCGAGTGATTTTCGGATTTGCGCCCTATCTTC 360
Db 5449 AAAGTGCTAAGGCTGTCGCAACCAAGTCTCGAGTGATTTTCGGATTTGCGCCCTATCTTC 5508
Qy 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 420
Db 5509 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 5568
Qy 421 TCTAGTGTACAGCATCATGTCGGCTGTTATCGGGGTTCTGCTCTTTCCACTTTTGGT 480
Db 5569 TCTAGTGTACAGCATCATGTCGGCTGTTATCGGGGTTCTGCTCTTTCCACTTTTGGT 5628
Qy 481 TTTATCTCTGAGTAACATATGACTCTCTCTCATCGCTATTATCTTAGCACCTTGTAGTTTG 540
Db 5629 TTTATCTCTGAGTAACATATGACTCTCTCTCATCGCTATTATCTTAGCACCTTGTAGTTTG 5688
Qy 541 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAATAAAGCTGAAAATTTGGTC 600
Db 5689 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAATAAAGCTGAAAATTTGGTC 5748
Qy 601 CCTTGGGATTTGAACCTAACCCATCAAGATCTCTAAAAAATAA 642
Db 5749 CCTTGGGATTTGAACCTAACCCATCAAGATCTCTAAAAAATAA 5790

RESULT 3

AAS55594

ID AAS55594 standard; DNA: 642 BP.

XX AC AAS55594;

XX DT 13-FEB-2002 (first entry)

XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #165.

XX KW Antisense; ds: prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Streptococcus pneumoniae.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX DR WPI; 2001-611495/70.

XX DR P-PSDB; AAU37735.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Claim 27; Seq ID No 9231; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, themselves and the discovery of novel antibiotics, the essential
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;

Query Match 99.8%; Score 640.4; DB 23; Length 642;

Best Local Similarity 99.8%; Pred. No. 1.1e-177;

Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGATTACAATAGTATTTTAAATCCTAGCCCTATCTGCTGGGTTTCGATTCATCTGGTCTC 60

Db 1 ATGATTACAATAGTATTTTAAATCCTAGCCCTATCTGCTGGGTTTCGATTCATCTGGTCTC 60

Qy 61 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCGAGCATGTTCTGGTAAACACTGGA 120

Db 61 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCGAGCATGTTCTGGTAAACACTGGA 120

Qy 121 ACACCAACACCTTCCCGCATTTTAGTAAAGAAAGCTGTATGGCAACCTTTGTGATTGAC 180

Db 121 ACACCAACACCTTCCCGCATTTTAGTAAAGAAAGCTGTATGGCAACCTTTGTGATTGAC 180

Qy 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTTATTTTCATCTACAAGGCGTTTCT 240

Db 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTTATTTTCATCTACAAGGCGTTTCT 240

Qy 241 CTTCTCATCTTTGGACTTTTGGCTGTTATCGCCATACCTTCCCTATCTTTGCGAGGATTT 300

Db 241 CTTCTCATCTTTGGACTTTTGGCTGTTATCGCCATACCTTCCCTATCTTTGCGAGGATTT 300

Qy 301 AAAGTGTAAGGCTGTCGCAACCAAGCTGCTGGAGTATTTTCGGATTTGCCCTATCTTC 360

Db 301 AAAGTGTAAGGCTGTCGCAACCAAGCTGCTGGAGTATTTTCGGATTTGCCCTATCTTC 360

Qy 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 420

Db 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 420

Qy 421 TCTAGTGTACAGCATCATGTTCCGCTGTTATCGGGTCTGCTCTTCCACTTTTGGT 480

Db 421 TCTAGTGTACAGCATCATGTTCCGCTGTTATCGGGTCTGCTCTTCCACTTTTGGT 480

Qy 481 TTTATCTCTGAGTAACATATGACTCTCTCTCATCGCTATTATCTTAGCACTTTGCTAGTTG 540

Db 481 TTTATCTCTGAGTAACATATGACTCTCTCTCATCGCTATTATCTTAGCACTTTGCTAGTTG 540

Qy 541 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAATAAAGCTGAAAATTTGGTC 600

Db 541 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAATAAAGCTGAAAATTTGGTC 600

Qy 601 CCTTGGGATTTGAACCTAACCCATCAAGATCTCTAAAAAATAA 642

Db 601 CCTTGGGATTTGAACCTAACCCATCAAGATCTCTAAAAAATAA 642

RESULT 4

```
AAS55801
ID AAS55801 standard; DNA; 642 BP.
XX
AC AAS55801;
XX
DT 13-FEB-2002 (first entry)
XX
DE Streptococcus pneumoniae DNA for cellular proliferation protein #372.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR P-PSDB; AAU37942.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 9438; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;

Query Match 99.8%; Score 640.4; DB 23; Length 642;
Best Local Similarity 99.8%; Pred. No. 1.le-177;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGATTACAAGTATTTTAACTACCTAGCTATCTGCTGGCTGCATTCCTCCATCTGGTCTC 60
DB 1 ATGATTACAAGTATTTTAACTACCTAGCTATCTGCTGGCTGCATTCCTCCATCTGGTCTC 60
OY 61 TGGATTGGACAAGTATTTTCAAAATCAATCTACCGAGCATGGTCTCTGCTAACACTGGA 120
DB 121 TGGATTGGACAAGTATTTTCAAAATCAATCTACCGAGCATGGTCTCTGCTAACACTGGA 120
OY 121 ACAGCAACACCTTCCGCATTTTAGGTAAAGAACTGGTATGGAACCTTTTGTGATTGAC 180
DB 121 ACAGCAACACCTTCCGCATTTTAGGTAAAGAACTGGTATGGAACCTTTTGTGATTGAC 180
OY 181 TTTTCAAGGAACCCCTAGCAACGCTCTCCGATTTATTTTTCATCTACAGGCGTTTCT 240
DB 181 TTTTCAAGGAACCCCTAGCAACGCTCTCCGATTTATTTTTCATCTACAGGCGTTTCT 240
OY 241 CCTCTCATCTTTGGACTTTTGGCTGTTTATCGGCCATACCTTCCCTATCTTTGCAGGATTT 300
DB 241 CCTCTCATCTTTGGACTTTTGGCTGTTTATCGGCCATACCTTCCCTATCTTTGCAGGATTT 300
OY 301 AAAGTGGTAAAGGCTGCGCAACGAGTCTGGAGTGAATTTTCGGATTTGGCGCTATCTTC 360
DB 301 AAAGTGGTAAAGGCTGCGCAACGAGTCTGGAGTGAATTTTCGGATTTGGCGCTATCTTC 360
OY 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTCTTTCATCGCTATCTTGGCAGTATGATTTCAC 420
DB 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTCTTTCATCGCTATCTTGGCAGTATGATTTCAC 420
OY 421 TCTAGTGTACAGCATCGATTGCGGCTGTTTATCGGGGTTCTGCTCTTTTCCACTTTTGGT 480
DB 421 TCTAGTGTACAGCATCGATTGCGGCTGTTTATCGGGGTTCTGCTCTTTTCCACTTTTGGT 480
OY 481 TTTATCTCTGAGTAACATGATGACTCTCTCTTTCATCGCTATCTTGGCAGTATGATTTCAC 540
DB 481 TTTATCTCTGAGTAACATGATGACTCTCTCTTTCATCGCTATCTTGGCAGTATGATTTCAC 540
OY 541 ATTATCATCTCGTATAAGGACAATATAGCTGATCAAAATATAAACTGAAATTTGGTC 600
DB 541 ATTATCATCTCGTATAAGGACAATATAGCTGATCAAAATATAAACTGAAATTTGGTC 600
OY 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAAA 642
DB 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAAA 642

RESULT 5
AAZ20352
ID AAZ20352 standard; DNA; 344 BP.
XX
AC AAZ20352;
XX
DT 17-NOV-1999 (first entry)
XX
DE Gene encoding bacterial general essential protein gep1493.
XX
KW General essential protein; pathogenic bacteria; pathogen; inhibitor;
KW bacterial growth; ds.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 2..343
FT /product= gep1493
XX
PN WO9933871-A2.
XX
PD 08-JUL-1999.
XX
PF 30-DEC-1998; 98WO-US27918.
XX
PR 31-DEC-1997; 97US-0070116.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Youngman P, Fritz C, Murphy C, Guzman L;
XX
DR WPI; 1999-430230/36.
DR P-PSDB; AAY22561.
```

Streptococcus pneumoniae general essential protein genes and proteins, useful for identification of antibacterial agents -

Claim 2; Fig 5; 124pp; English.

This sequence represents a Streptococcus pneumoniae general essential protein (GEP) gene of the invention. The genes encoding the GEP polypeptides are useful molecular tools for identifying similar genes in pathogenic microorganisms, such as pathogenic strains of Bacillus. In addition, the operons containing genes encoding GEP and the polypeptides themselves, are useful targets for identifying compounds that are inhibitors of the pathogens in which the GEP are expressed. Such inhibitors are useful for inhibiting bacterial growth by being bacteriostatic or bacteriocidal.

Sequence 344 BP; 73 A; 82 C; 71 G; 118 T; 0 other;

Query Match 51.1%; Score 327.8; DB 20; Length 344;
 Best Local Similarity 97.9%; Pred. No. 4e-86;
 Matches 332; Conservative 0; Mismatches 7; Indels 0; Gaps 0

QY 113 ACACGTGACGACGACACCTTCCGCATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTG 172
 |||||
 DB 6 ACATCGGAAGACCAACACCTTCCGCATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTG 65
 |||||
 QY 173 TGATTCACATTTTCAAAGGAACCTAGCAAGCGTCTCCGATTATTTTTCATCTACAAG 232
 |||||
 DB 66 TGATTCACATTTTCAAAGGAACCTAGCAAGCGTCTCCGATTATTTTTCATCTACAAG 125
 |||||
 QY 233 GCATTTCTCTCTCATCTTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTG 292
 |||||
 DB 126 GCGTTCTTCTCTCATCTTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTG 185
 |||||
 QY 293 CAGATTTAAAGTGTTAGGCTGTGCAACAGTGTCTGGAGTGATTTTCGGATTTGCGC 352
 |||||
 DB 186 CAGATTTAAAGTGTTAGGCTGTGCAACAGTGTCTGGAGTGATTTTCGGATTTGCGC 245
 |||||
 QY 353 CTATCTCTGTCTCTACCTTGGCATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGA 412
 |||||
 DB 246 CTATCTCTGTCTCTACCTTGGCATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGA 305
 |||||
 QY 413 TTTCAGTGTCTAGTGTACAGCATCGATTTGGCGTGTGA 451
 |||||
 DB 306 TTTCAGTGTCTAGTGTACAGCATCGATTTGGCGTGTGA 344
 |||||

RESULT 6
 ABN67449
 ID ABN67449 standard; DNA; 636 BP.
 XX ABN67449;
 AC
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Streptococcus polynucleotide SEQ ID NO 2811.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX

Db 481 TTTATATTAAAGGACTACGACTGATATTTACCACTGCTGTTATCTTAATGGCCCTAAACG 540
Qy 541 ATTATCATCTCGTATAGGACCAATATAGCTCGTATCAAAAATAAACTGAAATTTGGTC 600
Db 541 ATTATTCAGCATCAGGATAATATCAAACTGATTTGCAAAAAGGCAAGAAATTTGGTT 600
Qy 601 CCTTGGGATTGAACCTAAACCCATCAAGATCCCTAAA 636
Db 601 CCGTTGGCTTAAACCTAAGTAACAAAAAATAAA 636

RESULT 7
ABN70980
ID ABN70980 standard; DNA; 621 BP.
AC ABN70980;
DT 01-JUL-2002 (first entry)
DE Streptococcus polynucleotide SEQ ID NO 9873.
EX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS Streptococcus agalactiae.
PN WO200234771-A2.
PX 02-MAY-2002.
PY 29-OCT-2001; 2001WO-GB04789.
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettein H;
XX WPI; 2002-352536/38.
DR P-PSDB; ABP30349.
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein.
XX Claim 7; Page 4116; 4525pp; English.
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN56044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 621 BP; 172 A; 114 C; 104 G; 231 T; 0 other;
Query Match 43.0%; Score 276.2; DB 24; Length 621;

Best Local Similarity 65.5%; Pred. No. 7e-71;
Matches 404; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

Qy 20 TATATCCATGCTATCTGCTGGTTTCGATTCATCCATCTGCTCTGATTCGACAGTATTCT 79
Db 5 TATATCATCTGCTATCTGTTAGGCTCTATCCAACTGGACTTTGGATTGAAATATTTT 64
Qy 80 TCAAAATCAATCTACCGGAGCATGTTCTGTGAACACTGGAACGACCAACACCTCCGCA 139
Db 65 ACCAAGTCAATCTTCGTCACACGCTAGTGGTAATACTGACACCAATACCTCCGTA 124
Qy 140 TTTTAGGTAAGAAGCTGATGGAACCTTTTGTGATGACTTTTTCAAAGAACCTTAG 199
Db 125 TTTTAGGTGTAAGAAGCTGGCATTTGTCACATTTGACTATTGACATTTTAAAGGACTCTAG 184
Qy 200 CAACGCTGCTCCGATTATTTTTCATCTACAAGGGCTTTCTCTCTCATCTTTGGACTTT 259
Db 185 CAACACTTATTCCTATCATATTAGGTATACGACAGATCTCCATTTTATCGGTTTTT 244
Qy 260 TGGCTGTTATCGGCATACCTTCCCTATCTTTGCAAGGATTTAAAGTGTAAGGCTGTCG 319
Db 245 TTGCTATTATAGGTACACTTTTCCCTATTTTTCACAAATTCAAAGTGGAAGCGTTG 304
Qy 320 CAACAGTCTCGAGTGATTTTCGATTTGCGCCTATCTTCTGCTCTACCTTGGCATTA 379
Db 305 CTACAGCGCTGGGGTTTGGTAGGATTTGCACCTTCTTTTCTCTTACTTATTAGTTA 364
Qy 380 TCTTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTGCTAGTGTACAGCATCA 439
Db 365 TCTTTTATTAACGCTTATCTTTTTCAGTATTTTCCCTATCCAGTATTTACAGTAGCTG 424
Qy 440 TTGCGGCTGTTATCGGGTTCTGCTTTTCCACTTTTGGTTTATCTCTGACTAATATG 499
Db 425 TTGAGGTATCTTATGCTGCTTAAATTTTCCATTTAGTAGGCTTTTATATTACGGACTAG 484
Qy 500 ACTCTCTCTTCATCGCTATTATCTTTAGCAGCTTTGCTAGTTTGTATATCATCTCTCA 559
Db 485 ACTGGATATTTACCACCTGCTGTTATCTTAATGGCCCTTACAGGATTTATCCGACATCAG 544
Qy 560 ACAATATAGCTCTGATCAAAAATAAACTGAAATTTGCTCTCTGGGATTTGAACCTAA 619
Db 545 ATATATATCAAGGTATTCGAAAAGGCAAGAAATTTGGTCCGTTGGCTTAAACCTAA 604
Qy 620 CCCATCAAGATCCTAAA 636
Db 605 GTAAACAAAAATAAA 621

RESULT 8
ABN67450
ID ABN67450 standard; DNA; 720 BP.
XX AC ABN67450;
XX DT 01-JUL-2002 (first entry)
XX DE Streptococcus polynucleotide SEQ ID NO 2813.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS Streptococcus pyogenes.
XX
XX WO200234771-A2.
XX 02-MAY-2002.
XX 29-OCT-2001; 2001WO-GB04789.
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.

Db 5424 ATGAATAATCGTCATTTTGTACTTGTTCCTTATTTATAGTTCGATTCCCTCAGGTT 5365
 Oy 61 TGGATTGGACAAGTATCTTTCAATCAATACGAGCATGTTCTGTGAACACTGGA 120
 Db 5364 TGGATTGGTAACTTTCTTTTAAAGATATACGCCAATTTGGGAGTGGAAATACAGGA 5305
 Oy 121 ACGAACCAACCTTCCGCACTTTTAGTAAAGAAAGCTGGTATGGCAACCTTTGTGATGAC 180
 Db 5304 ACAACCAATACATTTCTGTCTTAGGAAACCTGCCGAATACGGTATTTAATGAT 5245
 Oy 181 TTTTCAAGGAACCTTAGCAAGCTCTTCCGATTTATTTTCAATPACAAAGCGTTCT 240
 Db 5244 ATCTTGAAGGAACGTTAGCCACTTCAATACCTATTTGTTGTTTACAAAGCGTGAAT 5185
 Oy 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTGAGGATTT 300
 Db 5184 CCGCTCTCTTTGGGTAGCAGCTGTTTAGGGCATACCTTCCCTATTTTGGCAATTC 5125
 Oy 301 AAGGTGGTAAGCGTGTGCAACAGCTGTGGAGTGTATTTTCGGATTTGCGCCTATCTTTC 360
 Db 5124 AAGGTGGTAAGCGGTAGCCACTAGCGCTGGCATGTTATACATACAGCCCATTT 5065
 Oy 361 TGCTCTACCTTCGGATTTCTTTTGGAGCTCTATCTTGGCAGTATGATTTCACTG 420
 Db 5064 TTTATTTATCTCTCTTTATTTTGTGATTTGTCTGTATCTAACCAGTATGGTGAGTTA 5005
 Oy 421 TCTAGTGTACAGCATCGATTCGGCTGTTATCGGGTGTCTGCTCTTCCACT-...TTTT 477
 Db 5004 ACAAGTATGATAGCCGTGTAATAATACACTTTCTACTATTATTTTGGCTTTCACTGTC 4945
 Oy 478 GGTTTTATCCTGAGTAACATGACTCTCTCTCATCGCTATTATCTTAGCACTTGTCTACT 537
 Db 4944 CCAGCTATTTTACCAACATTTAACTGGTTATTAACCGTATTTGCGGATTCCTTGACT 4885
 Oy 538 TTGATTTATCATGTCATAGGACAATATAGTCTGATCAAAATAAATACTGAAATTTG 597
 Db 4884 TTTATCTTTGTCGTCATCGTGAACATTCAGCGAATTAAGAAATGGAACAGAACTCGT 4825
 Oy 598 GTCCCTTGGGATTT 611
 Db 4824 CTCTCTTTGGCTT 4811

RESULT 10
 ID ABQ67195
 XX ABQ67195 standard; DNA; 495269 BP.
 AC ABQ67195;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria innocua contig DNA sequence #8.
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 KW infection; ds.
 OS Listeria innocua.
 XX
 PN W0200228891-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX

PT New genomic sequences from *Listeria* species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 XX
 PS Claim 5; SEQ ID 8; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (AB067198-AB067121) from *Listeria* sp. The sequences are useful as probes
 CC and primers for identification and/or detection of *Listeria* (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
 CC treating infections by *Listeria*, and are useful as immunogens in
 CC anti-*Listeria* vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 495269 BP; 159529 A; 88849 C; 93135 G; 153754 T; 2 other;

Query Match 19.3%; Score 124; DB 24; Length 495269;
 Best Local Similarity 56.1%; Pred. No. 4.8e-25;
 Matches 257; Conservative 0; Mismatches 195; Indels 6; Gaps 1;
 Oy 1 ATGATTACAAATAGTATTTTATTAATCCTAGCTATCTGCTGGTTCGATTCCATCTGCTC 60
 Db 441798 ATTAATTAATTTGCTTTCTTTGTTAGCTTATGTAATCGGTCAATACCTTTGCTTTA 441857
 Oy 61 TGGATTGGACAAGTATCTTTCAAAATCAATCTACGCGAGCATGGTCTGTAACACTGGA 120
 Db 441858 TGGATCGGTAAAAATTTTACAAAAAAGATATTCGTGATTTTGGTAGCGGAATTTAGGA 441917
 Oy 121 ACGACCAACACCTTCGCAATTTTAGTAGAAGAGCTGTATGGCAACCTTTCTGTATTCAC 180
 Db 441918 GCTACAAACTCCTTCGCTGTTTGGGGTAAAGCGGCAGTATTTGTTACAGTATGAT 441977
 Oy 181 TTTTCAAGGAACCTAGCAACGCTGCTTCGATTTATTTTCATCTACAAGCGTTCT 240
 Db 441978 ATTTTAAAGGTACTGTGCAACATTAATCTCCATCTTTTCCAACTAAATTTAATCAT 442037
 Oy 241 CCTCTCATCTT-----TGGACTTTTGGCTGTATTCGCGCCATACCTTCCCTATCTTTGCA 294
 Db 442038 CATTTCTGTTTACTTACGGGAGCATTTGCGATTTATCGGACATAGTTTCCCGCTTTTCGCA 442097
 Oy 295 GGATTTAAAGTGGTAAAGCTGTGCAACAGCTGCTGAGTGATTTTCGGATTTCCGCT 354
 Db 442098 GGATTCAGAGTGGGAAAGCAGTTGCTACTTCTGCTGTTATCTTCTTATGACCA 442157
 Oy 355 ATCTTCTCTCTACTTTCGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGAT 414
 Db 442158 CTTTGTGTTGAGCAGCACTAGTCTGTTTCTTATTATTAATAAATCAGCAATATG 442217
 Oy 415 TCATCTGCTAGTGTACAGCATGATTCGCTGTTTCTTATTAATAAATCAGCAATATG 452
 Db 442218 TCGCTTAGTTCATGATGGAGCAGCTTGCAGCATTAAT 442255

RESULT 11
 ABN92939
 ID ABN92939 standard; DNA; 615 BP.
 XX
 AC ABN92939;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2402.
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX

Db	487	TCATCATCATTCATCATGATATATATTTTACCTGTGTAGCGGAATTTGTTTCAATCATATTA 548
QY	517	ATATCTTACACATCTGCTAGTGTGATATATATCATTCGTC 554
Db	547	ATATTCGACACAAATCTAATATAGTTAGAAATTTTAA 584
RESULT 12		
AAV75059/c		
ID	AAV75059	standard; DNA; 818 BP.
XX	AAV75059;	
XX	16-MAR-1999	(first entry)
DT	XX	
DE	XX	Staphylococcus aureus contig SEQ ID #748.
XX	XX	Computer readable medium; vaccine; S.aureus infection; immunodetection
KW	XX	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy
KW	XX	skin infection; surgical wound infection; scalded skin syndrome;
KW	XX	toxic shock syndrome; ds.
OS	XX	Staphylococcus aureus.
XX	XX	
FH	Key	Location/Qualifiers
FT	misc_feature	121..180
FT		/tag= a
FT		/note= "these bases represent a line of missing text
FT		the sequence listing in the specification. The
FT		are included to maintain the nucleotide number
FT		given in the specification for this DNA sequence
XX	EP786519-A2.	
PN	XX	
XX	30-JUL-1997.	
PD	XX	
XX	07-JAN-1997;	97EP-0100117.
PF	XX	
XX	05-JAN-1996;	96US-0009861.
PR	XX	
XX	(HUMA-)	HUMAN GENOME SCI INC.
PA	XX	
XX	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;	
PI	Rosen CA;	
P1	XX	
XX	WPI; 1997-374922/35.	
DR	XX	
XX	Polynucleotide(s) and proteins derived from Staphylococcus aureus	
PT	stored on computer readable medium and used in the production of	
PT	anti-S.aureus vaccines	
XX		
PS	Claim 1; Page 1643-1644; 3271pp;	English.
XX	This sequence represents one of 5191 Staphylococcus aureus DNA sequences	
CC	of the invention. The DNA sequences are recorded on a computer readable	
CC	medium, preferably selected from a floppy or hard disk, random access	
CC	memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using	
CC	the S.aureus DNA sequences allows putative functions to be assigned	
CC	that protein-encoding or regulatory regions of commercial, therapeutic	
CC	industrial importance can be obtained. Specifically, sequences which	
CC	likely to encode antigens have been identified and these polypeptides	
CC	be used in a vaccine composition against S.aureus infection. The	
CC	polypeptides can also be used in a kit for the immunodetection of	
CC	S.aureus in a sample. S.aureus is implicated in numerous human diseases	
CC	including cellulitis, eyelid infections, food poisoning, osteomyelitis	
CC	skin and surgical wound infections, scalded skin syndrome, toxic shock	
CC	syndrome, etc. Organisms transformed with the DNA sequences can be used	
CC	for recombinant production of the polypeptides. The new DNA sequences	
CC	(and their fragments) are useful as primers or probes for isolating	
CC	homologues of any of the S.aureus DNA sequences contained on the	
CC	computer readable medium.	
XX	Sequence 818 BP; 294 A; 129 C; 112 G; 223 T; 60 other	
SQ		

Query Match 16.0%; Score 102.4; DB 18; Length 818;
 Best Local Similarity 52.6%; Pred. No. 7.2e-20;
 Matches 266; Conservative 0; Mismatches 216; Indels 24; Gaps 1;

QY 1 ATGATTACAAATAGTTTATTAATCCTAGCTATCTGCTGGGTTTCGATTCCCATCTGCTC 60
 DB 775 ATGATGATAATCGTCATGTTACTACTAATGTTATCTATCTGCGGCTTCCCAAGTGGATTC 716
 QY 61 TCGATTGGACAAGTATCTTCAAAATCAATCTACCGGAGCATGGTCTGGTAACACTGGA 120
 DB 715 GTAATGGGAATATTTTCAAAAAGATATTAGACAATTTGGTAGTGGTAATACTGGC 656
 QY 121 AGGACCAACACTTCCCGCATTTTAGTAAAGAAAGCTGTATGGCAACCTTTGTGATTCAC 180
 DB 655 GCTACTAATAGCTTTAGAGTATTAGGTGCTCTCGAGGATTTCTGGTAACATTTCTAGAT 596
 QY 181 TTTTCAAGGACCTAGCAACGCTGCTCCGATTTATTTTCTATCTACAAAGCGGTTTCT 240
 DB 595 ATTTTCAAGGCTTCAAACTGTTTCTTCCCTTTATGTTTACAAAGCTTTCAGCAGATGGC 536
 QY 241 CCTCTC-----ATCTTGGACTTTTGGCTGTTATCGGCCAT 276
 DB 535 CCTATAGTACTTTTTTACAAATGGTTTAAATCTGGCTTATTCGCTACTTGGACAC 476
 QY 277 ACCTTCCCTATCTTTTGGAGGATTTAAAGGTGTAAGGCTGTGCAACAGCTGGAGTG 336
 DB 475 GTTATCTCTGTTTATTTAAATTTCAAGGTGGCAAGCTGTGCAACTAGTGCAGGTGC 416
 QY 337 ATTTTCCGATTTGGCCCTATCTCTCTACCTTCCGATTTATCTTCTTTTGGAGCTCTC 396
 DB 415 GTCTTGGAGTCAATCCGATACTTTTACTAATCTTCAATTTCTTTTATTTGATTTG 356
 QY 397 TATCTTGGCAGTATGATTTTCACTGCTAGTGTGCACAGCATCGATGCGGCTGTTATCGGG 456
 DB 355 AAGATTTTAAATATGTTTCTTTAGCAAGTATCGTTGCAAGATTTGCTGTGATTTGGC 296
 QY 457 GTTCTGCTTTTCCACTTTTTTGGTTT 482
 DB 295 TCGCTTATCTTCAAGACTATATTT 270

RESULT 13
 AH54476/c
 ID AH54476 standard; DNA; 3014 BP.
 AC AH54476;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3840.
 XX
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 useful for vaccinating against infections, e.g. endocarditis -

XX
 PS
 XX Claim 8; Page 1482-1483; 2188pp; English.
 CC AH52304 to AH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AH53970 to AH53120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AH53971 to
 CC AH55098 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AH55091 to
 CC AH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 3014 BP; 1083 A; 444 C; 510 G; 977 T; 0 other;
 Query Match 16.0%; Score 102.4; DB 22; Length 3014;
 Best Local Similarity 50.7%; Pred. No. 1.2e-19;
 Matches 233; Conservative 0; Mismatches 261; Indels 24; Gaps 1;

QY 1 ATGATTACAAATAGTTTATTAATCCTAGCTATCTGCTGGGTTTCGATTCCCATCTGCTC 60
 DB 1223 ATGATGATCATCTCATGTTTAACTTCTGAGTTATCTGATGGTCATTCACAGCGGTTA 1164
 QY 61 TCGATTGGACAAGTATCTTCAAAATCAATCTACGCGAGCATGGTCTGTGTAACACTGGA 120
 DB 1163 ATTTTGGTAAATATTTTAAAAAAGATATAAGACAATACGGTAGTGAATACTGGA 1104
 QY 121 AGCACCACACCTTCCGCTTTTAGTGAAGAGCTGGTATGGAACACCTTTGTGATTCAC 180
 DB 1103 GCACTAACAGTTTCTGCTTCTTGAAGACAGCTGGATTTATAGTTTACGTTTATAGAT 1044
 QY 181 TTTTCAAGGAAACCTTACGACGCTTCCGATTTATTTTCACTACAAGCGTT--- 237
 DB 1043 ATTTTCAAGGATTTTATACAGCTTTTTCACATATGTTTCCAGTCTCAGGATGGT 984
 QY 238 -----TCTCCTCTCATCTTTGGACATTTGGCTGTTATCGGCCAT 276
 DB 983 GTTATAAGCACCTTCTTTTACAAATGTTTAAATAGTAGGATTTGTCATCTCGGTAC 924
 QY 277 ACCITCCCTATCTTTGAGGATTTAAAGGTGTTAGGCTGTCGCAACAGCTGCTGAGTG 336
 DB 923 GTGATCCAAATATATCTGAAATTTAATGGCGAAAGCAGTAGCTACAGTGCAGGAGTT 864
 QY 337 ATTTTGGGATTTGCGCTTATCTTCTCTACCTTGGATTTGCTGTTATCGGCTCTC 396
 DB 863 GTATTAGGTGTCATCTTATTTACTTCTATCTTGGCAATATCTTTTATAGTATTA 804
 QY 397 TATCTTGGCAGTATGATTTTCACTGCTGTAGTGTACAGCATCGATTTGCGCTGTTATCGG 456
 DB 803 AAAATCTTTAAATATATGTTTCTTATCAAGTATCATTCAGCAATAGTTTGTGATGGT 744
 QY 457 GTTCTGCTTTTCCACTTTTGGTGTATCTGCTAGTAACTAGTCTCTCTTCATCGCT 516
 DB 743 TCAATCATCATTCATGATTTATTTTACTTGTCTGTAGCGGAATTTTCAATCATATTA 684
 QY 517 ATTTTCTTAGCACCTTGTAGTTTCTGATTTATCATTCGTCAC 554
 DB 683 ATAATTCGACACAAATCTAATATAGTTAGATTTTTTAA 646

RESULT 14
 AH54115/c

CC themselves, are useful targets for identifying compounds that are
CC inhibitors of the pathogens in which the GEP are expressed. Such
CC inhibitors are useful for inhibiting bacterial growth by being
CC bacteriostatic or bacteriocidal.
xx

SQ Sequence 582 BP; 147 A; 115 C; 122 G; 198 T; 0 other;

```
Query Match      15.6%; Score 100; DB 20; Length 582;
Best Local Similarity 52.9%; Pred. No. 3.2e-19;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 ATGATTACAAATAGTTTATTAATCCTAGCTATCTGCTGGGTTTCGATCCATCTGGTCTC 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1 ATGTTAAATGCTTTATTTGATTTATTTGGCTACTTTGATAGGCGAGCATTCCTCTGGCTTA 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 TGGATTGGACAGATATCTTTCAAATCAATCTACCGGAGCATGGTCTGGTAACACTGGA 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 ATTGTGGCAAGCTTGGCAAGGAATGATATTCGGGAGCACGGAAGCGCAACTTAGGC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 ACGACCAACACCTCCGCAATTTAGGTAAGAAAGCTGCTATGGCAACCTTTGTGATTGAC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 GCTACCAATGCATTCGCTACATGGGTGTAAGAGCTGGTTCGGTGCATAGCCGGAGAT 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 TTTTCAAGGAACCTAGCAACGCTGCTCCGATTTATTTTCATCTACAAGGCGTTTCT 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 ATTTGAAAGGACACTGGCAACTGCATTGCTTTTCTCATGCATGT---TGATATTCAC 237
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 CCTCTCATCTTTGGACITTTGGCTGTATCGGCCATACCTTCCCTATCTTTGCAGGATTT 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 238 CCGCTTCTTGCGAGGATCTTTGGCGTTTTAGGCCACCGTGTTCCTCATCTTCGCCCAATTT 297
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 AAAGTGTAAGGCTGTCGCAACCAAGTCTGCTGGAGTATTTTCGGATTTGCGCCTATCTTC 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 298 AAAGCGGTAAGCGGTGGCGACATCAGGAGCGCTTTTGCTATTTTACGCACCCCTGTTA 357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 TGCTCTACCTTGGGATATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 358 TTTATCAGATGGTTGGGGTATCTTCTATCTTTTATACTTGACTAAATTTGTTTCTCTC 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 TCTAGTGTACAGCATCGATTCGGCTGTTAT 452
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 418 TCATCGATGTTACAGGGATCTATACTGTTAT 449
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: December 24, 2002, 22:43:13
Job time : 497.797 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:16:24 ; Search time 3529.69 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estinu:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rpd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	51.6	8.0	997	17	CNS005TE
c	2	44.6	6.9	1154	12 BF689161
c	3	42.6	6.6	634	9 AL513733
c	4	42.4	6.6	1267	12 BG114247
c	5	42	6.5	633	17 BH398276
c	6	42	6.5	772	17 BH384505
					AL060767 Drosophil
					BF689161 602184801
					AL513733 AL513733
					BG114247 602285812
					BH398276 AG-ND-138
					BH384505 AG-ND-138

c	7	42	6.5	1058	17	AG056521
c	8	41.6	6.5	670	13	BJ153618
c	9	41.2	6.4	500	9	AU086536
c	10	41.2	6.4	1101	17	CNS00399
c	11	41	6.4	827	17	CNS02156
c	12	41	6.4	838	17	CNS03W9J
c	13	41	6.4	967	17	CNS0011A
c	14	41	6.4	1015	17	CNS00H5T
c	15	40.8	6.4	796	17	AG063366
c	16	40.8	6.4	1101	17	CNS00240
c	17	40.6	6.3	868	17	AZ547264
c	18	40.6	6.3	876	17	AZ529709
c	19	40.6	6.3	877	17	AZ547883
c	20	40.6	6.3	890	17	BH135953
c	21	40.6	6.3	899	17	BH147716
c	22	40.6	6.3	957	17	AG051372
c	23	40.4	6.3	500	9	AU088092
c	24	40.4	6.3	544	13	BJ420171
c	25	40.4	6.3	1101	17	CNS00182
c	26	40.2	6.3	588	9	AA541617
c	27	40.2	6.3	789	9	AI658642
c	28	40.2	6.3	821	17	B21581
c	29	40.2	6.3	821	17	AG031754
c	30	39.8	6.2	421	13	BI670700
c	31	39.8	6.2	430	13	BI815236
c	32	39.8	6.2	503	9	AL513809
c	33	39.8	6.2	730	12	BE866465
c	34	39.8	6.2	745	14	BQ986791
c	35	39.6	6.2	344	9	AL513975
c	36	39.6	6.2	709	17	BH402216
c	37	39.4	6.1	1089	12	BG476766
c	38	39.4	6.1	1101	17	CNS00KK2
c	39	39.2	6.1	329	9	AL513719
c	40	39.2	6.1	1058	17	CNS011HR
c	41	39	6.1	987	17	CNS00418
c	42	38.8	6.0	408	13	BM160537
c	43	38.8	6.0	431	17	CNS04J60
c	44	38.8	6.0	504	9	AL514489
c	45	38.8	6.0	510	13	BM162029

ALIGNMENTS

RESULT 1	CNS005TE	CNS005TE	997 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	Drosophila melanogaster	Drosophila melanogaster	genome survey	sequence	TET3 end of BAC #	
DEFINITION	BAC12K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	BAC12K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL060767	AL060767.1	GI:4943573			
VERSION	GSS.					
KEYWORDS	Drosophila melanogaster.					
SOURCE	Drosophila melanogaster.					
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 1 to 997)					
AUTHORS	Genoscope.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)					
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Anron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial					

Db 711 TTCCTTTCTGTGTTTTCTTTTTTTTTTCCCTCTCNCCTTCTCTTTTTCTTTTTCTCTCT 652

<hr/>					
RESULT 9	AU086536	500 bp	mRNA	linear	EST 27-JAN-2001
LOCUS	AU086536 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA				
DEFINITION	clone XPFn2783, Malaria sequence.				
ACCESSION	AU086536				
VERSION	AU086536.1	GI:12388677			
KEYWORDS	EST.				
SOURCE	Plasmodium falciparum 3D7.				
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
REFERENCE	1 (bases 1 to 500)				
AUTHORS	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.				
TITLE	PULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum				
JOURNAL	Nucleic Acids Res. 29 (1), 70-71 (2001)				
MEDLINE	20574754				
COMMENT	Contact: Junichi Watanabe Institute of Medical Science The University of Tokyo, Department of Parasitology 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Tel: 81-3-5449-5378 Fax: 81-3-5449-5410 Email: jwatanab@manage.ims.u-tokyo.ac.jp Suzuki,Y., Yoshimoto-Nakagawa,K., Haruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).				
FEATURES	Location/Qualifiers				
source	1..500				
	/organism="Plasmodium falciparum 3D7"				
	/strain="3D7"				
	/db_xref="taxon:36329"				
	/clone="XPFn2783"				
	/clone_lib="Sugano Malaria cDNA library"				
	/dev_stage="erythrocytic stage"				
BASE COUNT	278 a 43 C 34 g 142 t 3 others				
ORIGIN					
Query Match	6.4%; Score 41.2; DB 9; Length 500;				
Best Local Similarity	47.3%; Pred. No.1.2;				
Matches	124; Conservative 0; Mismatches 138; Indels 0; Gaps 0;				
Qy	335 TGATTTTCGGATTTCGGCTATCTCTCTCTACCTTGCAGTATTATCTCTTTGGAGCTC	394			
Db	450 TTAATTATTGTTACTAATTATGTGAATAATTATGTTGACTAATTATTATTATTATTGTTA	391			
Qy	395 TCTATCTTGGCAGTAGTTTCACTGCTAGTGTCACAGCATCGATTGCGGCTGTTATCG	454			
Db	390 TTATTGTTATTATTATTATTATTGTTGTTACTATTATTATTATTGTTACTATTATTA	331			
Qy	455 GGGTTCCTCTCTTCCACCTTTTGGTTTTATCCTCGAGTAATGACATGCTCTCTTCATCG	514			
Db	330 TTAATTATTATTATTATTATTGTTGTTGTTACTATTATTATTGTTGTTACTATTATTG	271			
Qy	515 CTATTATCTAGCACTTCCTAGTTTGGATTATTCATTTCGTCATAAGCACAAATATAGCTCGTA	574			
Db	270 TTGCTATTATTATTGTTGCTATTATTATTATTATTATTATTATTATTATTATTATTATTA	211			
Qy	575 TCAGAAAATAAACCTGAAAAATTT 596				
Db	210 TTAATTATTATTGTTATTAT 189				
RESULT 10	CNS00399/c				
LOCUS	CNS00399 Drosophila melanogaster genome survey sequence TEF3 end of BAC #				
DEFINITION	BACR08E12 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL063914				
VERSION	GSS.3914				
KEYWORDS	GSS.				

063D18 of library G from Tetraodon nigroviridis, genomic survey sequence.										
ACCESSION	AL263440									
VERSION	GI:7985102									
KEYWORDS	GSS: genome survey sequence.									
SOURCE	Tetraodon nigroviridis.									
ORGANISM	Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 838) Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence Unpublished 2 (bases 1 to 838) Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier.F., Saurin,W., Bernot.A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Unpublished 3 (bases 1 to 838) Direct Submission Submitted (12-APR-2000) This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon. 									
TITLE										
JOURNAL										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
FEATURES	Location/Qualifiers 1..838 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="063D18" /clone_lib="G" /note="Genoscope sequence ID : COBG063DB09SP1-end : PUC-Ori"									
BASE COUNT	97	a	110	c	18	g	490	t	123	others
ORIGIN										
Query Match	6.4%; Score 41; DB 17; Length 838;									
Best Local Similarity	35.1%; Pred. No. 1.4;									
Matches	140;	Conservative	47;	Mismatches	212;	Indels	0;	Gaps		
QY	206	TGCTTCGGATATATTTTCATCFACAGCGGTTCCTCCTCATCTTTGGACTTTTGCGTG	265							
Db	290	TT	349							
QY	266	TTATCGGCATACCTCCCTATCTTTGCAGGATTAAAGGTGGTAAGCGCTGCCAACCA	325							
Db	350	TATTTTCCCTCTTT	409							
QY	326	GTGCTGGAGTGATTTTCGGATTTCGCCCTATCTTCCTGCTCTACTCGCATATATCTTCT	385							
Db	410	TTCTTTTTCCTTTTCTCTCTTTTYYTYTYTTCCTTTTTTTTTTTTTTTTTTTTTTTT	469							
QY	386	TTGGAGCTCTCTATCTTGGCAGTATGATTTCACTCTCTAGTGTCAACAGCATCGATCGGG	445							
Db	470	TTCTTTTTTCTTTTTTTTTTTTTTCTTKTKKTKTKTCTTCTCYKTGCTKKKKKCTKK	529							
QY	446	CTGTTATCGGGTTCGCTCTTTCCACATTTTGGTTTTTATCCTGAGTAGTATGACTCTC	505							
Db	530	TTTK	589							
QY	506	TCTTCATCGCTATTATCTTAGCACCTTGCTAGTTCATTATCATTCGTCATAAGGAACA	565							
Db	590	TCHTTK	649							
QY	566	TAGCTCGTATCAAATAAAAACTGAAATTTTCGGTCCCCTT	604							

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BAC34J03 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL0731171

VERSION AL0731171.1

KEYWORDS GI:4952850

SOURCE GSS.

ORGANISM Drosophila melanogaster.

Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1015)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source 1..1015

1..organism="Drosophila melanogaster"

1..db_xref="taxon:7227"

1..clone="BACR34J03"

1..clone.lib="RPCI-98"

1..note="end : T7"

BASE COUNT 280 a 142 c 199 g 245 t 149 others

ORIGIN

Query Match 6.4%; Score 41; DB 17: Length 1015;

Best Local Similarity 48.7%; Pred. No. 1.4;

Matches 93; Conservative 6; Mismatches 92; Indels 0; Gaps 0;

229 GCGGTTTCCTCCTCATCTTTCGACATTTTGGCTGTTATCGGCATACCTTCCTATCTTT 291

230 GGCATTTTTCCTCCTGTTTTTTCGCTTTTTCCTCCTTTTTCCTCCTGCTTTT 144

231 GCAGGATTTAAAGGTGAAGCTCGCAACAGCTGCTGGAGTATTTTCGGATTTGGC 351

232 TTTCCCTTACACGGTTKTTTGGTTTTCATTTATTTTGTAGTGGGTTTGGTTTGGG 84

233 CTTATCTTCTGCTCTACCTTCGCGATTTCTTTTGGAGCTCTCTATCTTTCGAGTATG 411

234 TTTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

235 ATTTCACTGTC 422

236 TTTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

237 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

238 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

239 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

240 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

241 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

242 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

243 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

244 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

245 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

246 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

247 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

248 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

249 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

250 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

251 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

252 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

253 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

254 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

255 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

256 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

257 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

258 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

259 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

260 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

261 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

262 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

263 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

264 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

265 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

266 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

267 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

268 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

269 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

270 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

271 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

272 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

273 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

274 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

275 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

276 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

277 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

278 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

279 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

280 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

281 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

282 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

283 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

284 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

285 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

286 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

287 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

288 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

289 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

290 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

291 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

292 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

293 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

294 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

295 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

296 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

297 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

298 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

299 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

300 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

301 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

302 ATTTTCCTTTCGTTGCGVGTGTTTTTAAABACTAATTTGGCGGCGATGTTCTCGCATGCC 24

303 ATTTTCCTTTCGTTGCGVGTGTTTT

[illegible]

Search completed: December 25, 2002, 14:17:10
Job time : 3536.69 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:25:54 : Search time 93.625 Seconds
(without alignments)
2102.926 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	642	100.0	6171	4	US-08-961-527-37
2	327.8	51.1	344	4	US-09-222-938A-14
3	102.4	16.0	615	4	US-09-134-001C-2402
4	100	15.6	582	4	US-09-222-938A-71
5	66	10.3	7218	1	US-08-232-463-14
6	34	5.3	521	4	US-09-221-017B-530
7	33.6	5.2	630	1	US-08-375-241-10
8	33.6	5.2	630	5	PCT-US92-06617A-10
9	31.8	5.0	2439	4	US-09-632-098-6
10	31.8	5.0	3531	2	US-08-714-402-1
11	31.6	4.9	277	4	US-09-007-005-3
12	31.6	4.9	277	4	US-09-244-796-3
13	31.6	4.9	5150	4	US-09-068-140A-14
14	31.2	4.9	99500	4	US-09-798-096-10
15	31.2	4.9	169998	4	US-09-676-610B-24
16	31	4.8	248	4	US-09-007-005-32
17	31	4.8	248	4	US-09-244-796-32
18	31	4.8	557	3	US-08-776-660A-1
19	31	4.8	624	3	US-08-776-660A-2
20	30.8	4.8	72604	4	US-09-268-992-7
21	30.8	4.8	72604	4	US-09-657-474-7
22	30.6	4.8	637	4	US-09-328-111-628
23	30.6	4.8	1074	2	US-08-975-316-54
24	30.6	4.8	1074	4	US-09-615-192A-54
25	30.6	4.8	1075	2	US-08-975-316-55
26	30.6	4.8	1075	4	US-09-615-192A-55
27	30.4	4.7	1924	4	US-08-687-590-55

C 28	30.4	4.7	44453	4	US-09-146-053-5	Sequence 5, Appli
C 29	30.2	4.7	289	4	US-09-007-005-17	Sequence 17, Appl
C 30	30.2	4.7	289	4	US-09-244-796-17	Sequence 17, Appl
C 31	30.2	4.7	2216	1	US-08-046-583-12	Sequence 12, Appl
C 32	30.2	4.7	2216	1	US-08-384-556A-3	Sequence 3, Appli
C 33	30.2	4.7	2216	2	US-08-331-355A-24	Sequence 24, Appl
C 34	30.2	4.7	2216	5	PCT-US94-1236A-24	Sequence 24, Appl
C 35	30.2	4.7	2216	5	PCT-US95-07753-3	Sequence 3, Appli
C 36	30.2	4.7	2298	3	US-09-157-077-7	Sequence 7, Appli
C 37	30.2	4.7	2327	3	US-09-157-077-1	Sequence 1, Appli
C 38	30	4.7	4248	3	US-08-678-614-1	Sequence 1, Appli
C 39	30	4.7	9848	4	US-09-385-222A-3	Sequence 3, Appli
C 40	30	4.7	11340	4	US-08-961-527-147	Sequence 147, App
C 41	29.6	4.6	451	4	US-09-404-879A-103	Sequence 103, App
C 42	29.6	4.6	1958	4	US-08-961-527-360	Sequence 360, App
C 43	29.6	4.6	2186	4	US-08-961-527-179	Sequence 179, App
C 44	29.6	4.6	2651	4	US-08-961-527-216	Sequence 216, App
C 45	29.6	4.6	2674	2	US-08-926-724-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-961-527-37
; Sequence 37, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-37

Query Match 100.0%; Score 642; DB 4; Length 6171;
Best Local Similarity 100.0%; Pred. No. 7.4e-193;
Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGATTACAATAGTTTATTAACTAGCCTATCTGCTGGTTGCGATTCATCGGTC 60
DB 5149 ATGATTACAATAGTTTATTAACTAGCCTATCTGCTGGTTGCGATTCATCGGTC 5208

```
QY 61 TGGATTGACAAAGTATTTCTTTCAAAATCAATCTACCGGAGCATGGTTCTGGTAAACACTGGA 120
DB 5209 TGGATTGACAAAGTATTTCTTTCAAAATCAATCTACCGGAGCATGGTTCTGGTAAACACTGGA 5268
QY 121 ACGACCAACACTTCCGCAATTTAGGTAAAGAACTGGTATGGCAACCTTTGTGATTGAC 180
DB 5269 ACGACCAACACTTCCGCAATTTAGGTAAAGAACTGGTATGGCAACCTTTGTGATTGAC 5328
QY 181 TTTTTCAAAGAACCTTAGCAACGCTGCTTCCGATTATTTTTCATCTACAAGGCGTTTCT 240
DB 5329 TTTTTCAAAGAACCTTAGCAACGCTGCTTCCGATTATTTTTCATCTACAAGGCGTTTCT 5388
QY 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
DB 5389 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 5448
QY 301 AAAGTGCTAAGGCTGTCGCAACCACTGCTGGAGTGATTTTTCGGATTTCGGCTTATCTTC 360
DB 5449 AAAGTGCTAAGGCTGTCGCAACCACTGCTGGAGTGATTTTTCGGATTTCGGCTTATCTTC 5508
QY 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
DB 5509 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 5568
QY 421 TCTAGTGTACAGCATGCAATTCGGCTGTTATCGGGTTCGCTCTTCCACTTTTGGT 480
DB 5569 TCTAGTGTACAGCATGCAATTCGGCTGTTATCGGGTTCGCTCTTCCACTTTTGGT 5628
QY 481 TTTATCTCTGAGTAAGTATGCTCTCTCTTCATCGCTATTATCTTAGCACTTGTAGTTTG 540
DB 5629 TTTATCTCTGAGTAAGTATGCTCTCTCTTCATCGCTATTATCTTAGCACTTGTAGTTTG 5688
QY 541 ATTATCATCTGCTATAAGGACAATATAGCTCGTATCAAAAAATAAACTGAAAAATTTGGTC 600
DB 5689 ATTATCATCTGCTATAAGGACAATATAGCTCGTATCAAAAAATAAACTGAAAAATTTGGTC 5748
QY 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAAATAA 642
DB 5749 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAAATAA 5790
```

```
RESULT 2
US-09-222-938A-14
; Sequence 14, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 344
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(343)
US-09-222-938A-14

Query Match 51.18; Score 327.8; DB 4; Length 344;
Best Local Similarity 97.98; Pred. No. 3.3e-94;
Matches 332; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
QY 113 ACATCTGGAACCAACCACTTCCGCAATTTAGGTAAAGAACTGGTATGGCAACCTTTG 172
DB 6 ACATCTGGAACCAACCACTTCCGCAATTTAGGTAAAGAACTGGTATGGCAACCTTTG 65
```

```
QY 173 TGATTGACTTTTTCAAAGGAACCCCTAGCAAGCGTGTCTCCGATTATTTTTCATCTACAAG 232
DB 66 TGATTGACTTTTTCAAAGGAACCCCTAGCAAGCGTGTCTCCGATTATTTTTCATCTACAAG 125
QY 233 GCSTTTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTG 292
DB 126 GCSTTTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTG 185
QY 293 CAGGATTTAAAGGTGGTAAAGGCTCTCGCAACCACTGCTGGAGTGATTTTTCGGATTTGCGC 352
DB 186 CAGGATTTAAAGGTGGTAAAGGCTCTCGCAACCACTGCTGGAGTGATTTTTCGGATTTGCGC 245
QY 353 CTATCTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCTTTGGAGCTCTCTATCTTTGGCAGTATGA 412
DB 246 CTATCTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCTTTGGAGCTCTCTATCTTTGGCAGTATGA 305
QY 413 TTTCACTGCTCTAGTGTACAGCATCGATTCGGCTGTTA 451
DB 306 TTTCACTGCTCTAGTGTACAGCATCGATTCGGCTGTTA 344
```

```
RESULT 3
US-09-134-001C-2402
; Sequence 2402, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2402
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2402
```

```
Query Match 16.0%; Score 102.4; DB 4; Length 615;
Best Local Similarity 50.7%; Pred. No. 7.8e-23;
Matches 293; Conservative 0; Mismatches 261; Indels 24; Gaps 1;
```

```
QY 1 ATGATTACAAATAGTATTTTAAATCCTAGCCTATCTGCTGGTTCGATTCCATCTCGTCTC 60
DB 7 ATGATTACAAATAGTATTTTAAATCCTAGCCTATCTGCTGGTTCGATTCCATCTCGTCTC 66
QY 61 TGGATTGCAACAAGTATTTCTTTCAAAATCAATCTACGCCAGCATGGTCTCTGTAACACTGGA 120
DB 67 ATGATTGCAACAAGTATTTCTTTCAAAATCAATCTACGCCAGCATGGTCTCTGTAACACTGGA 126
QY 121 ACGACCAACACTTCCGCAATTTAGGTAAAGAACTGGTATGGCAACCTTTGTGATTGAC 180
DB 127 GCAACTAACAGTTTTCTGTTCTTTGGAAGAACCACTGCGATTATAGTTAGCTTTTATGAT 186
QY 181 TTTTTCAAAGGAACCCCTAGCAACGCTGCTTTCCGATTATTTTTCATCTACAAGGCGTT--- 237
DB 187 ATTTTCAAGGATTTTATACAGTCTTTTTCACACTGTTTTCACACTGTTTCCAGTTTCAGCGGAT 246
QY 238 -----TCTCCTCTCATCTTTTGGACTTTTGGCTGTTTATCTGATCTATCTGCGCAT 276
DB 247 GTTATAAGCACTTCTTTTCAAAATGGTTTAAATAGTAGGATGTTTGGCAACTACTCGGTAC 306
QY 277 ACCCTCCCTATCTTTGAGGATTTAAAGGTTAAAGGCTGTGCAACCACTGCTGAGTG 336
DB 307 GTGTATCCAATATATCTGAAATTTAAATGCGGAAAGCACTAGCTACCACTGCGAGGAT 366
QY 337 ATTTTCCGATTTGCGCTATCTCTGCTCTACCTTCCGATTATCTTCTTTGGAGCTCTC 396
```

Db 367 GTATTAGGTGCTCAATCTATTTTACTTCTTATCTTGCAATATATCTTTTACTGTATTA 426
Qy 397 TATCTGGCAGTAGATTTTACCTAGTGTCTACAGCATGATTGGCGCTGTATTCGGG 456
Db 427 AAAATCTTTAAATATGTTCTTTTATCAAGTATCATTCGACGAATAGTTGTGTGATGGT 486
Qy 457 GTTCTGCTCTTTCCACCTTTTGTGTTTATCCTCAGTAACATGACATCTCTCTTCATCGCT 516
Db 487 TCAATCATCATCATGATATATATTTTACTTGTCTGTAGCGAATGTTTCAATCATATTA 546
Qy 517 ATTATCTAGCAGTGTGCTAGTGTGATGATATCATTCGTCA 554
Db 547 ATAATCCACACAATCTAATATAGTATAGAAATTTTAA 584

RESULT 4
US-09-222-938A-71
; Sequence 71, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritze, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 71
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(579)
US-09-222-938A-71

Query Match 15.6%; Score 100; DB 4; Length 582;

Best Local Similarity 52.9%; Pred. No. 4.4e-22;

Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 ATGATTACAATGTTTATTAATCCTAGCTATCTGCTGGTTCGATTCATCTGGTCTC 60
Db 1 ATGTTAATGCTTTATGATTTATTTGGCTACTGTATGAGCAGCATTCCTCTGGCTTA 60
Qy 61 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCCAGCATGTTCTGCTAACACTGGA 120
Db 61 ATTGTGGCAGCTTGCCAAAGGAATGATATTCGGGAGCAGCAGCGCACTTAGGC 120
Qy 121 ACAGCAACACCTTCCGCAATTTAGGTAGAAAGCTGGTATGGCAACCTTTGTGATGAC 180
Db 121 GCTACCAATGATTCCTGATATTTGGTGTAAAGCTGGTTCGCTGATAGCGGAGAT 180
Qy 181 TTTTCAAGCAACCTTAGCAACCTCTGCTCCGATTTATTTTCTATCTACAGGCGTTCT 240
Db 181 ATTTTGAAGGACCTAGGCACTGCAATGCTTTCTCATGATG---TGATATTCAC 237
Qy 241 CCTCTCATCTTTTGACCTTTTGGCTGTTATGGCCATACCTTCCCTATCTTTGCGAGATT 300
Db 238 CCGCTCTTGGCAGAGCTTTGCGGTTTATAGCCACCTGTTTCCCATCTTCGCAATTT 297
Qy 301 AAAGTGGAAGGCTGCGCAACAGTGGTGGAGTATTTTGGATTTGGCCCTATCTTC 360
Db 298 AAAGCGGTAAGCGGTGGGACATCAGAGGCGTTTGTATTTTACGCAACCCCTGTA 357
Qy 361 TGCTCTACCTTGGATATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
Db 358 TTTATCAGATGGTGGGATTTCTTATCTTTTATCTGACTAAATTTGTTTCTCTC 417
Qy 421 TCTAGTGCACAGCATGATTCGGCTGTTAT 452

Db 418 TCATCGATCTTAACAGGATCTATCTACTGTAT 449

RESULT 5

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/935,313

; FILING DATE:

; APPLICATION NUMBER: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: pTZ9pt-Fls

US-08-232-463-14

Query Match 10.3%; Score 66; DB 1; Length 7218;

Best Local Similarity 3.1%; Pred. No. 9e-11;

Matches 12; Conservative 231; Mismatches 141; Indels 0; Gaps 0;

Qy 196 CTAGCAACGCTGCTCCGATTTATTTTCTCATCTACAGGCTTCTCTCTCATCTGTTGGA 255
Db 1059 CTTCGATT 1118
Qy 256 CTTTTGGCTTTATCGCCATACCTTCCCTATCTTTGCAGGATTTAAAGTGGTAAGGCT 315
Db 1119 YY 1178
Qy 316 GTCCCAACCACTGCTGAGTGAATTTTCGGATTTGCCCTATCTTCTCTACCTTGGC 375
Db 1179 YY 1238
Qy 376 ATTATCTTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTGCTAGTGTACAGCA 435
Db 1239 YY 1298
Qy 436 TCGATCGGCTGTTATCGGGTCTCTCTCTTCCACTTTTGGTTTTTATCTCTGAGTAAC 495

Db 289 TTTTCAACATTAGTTTAACTTCTAGTTTATTTCTTATATCTTCCATCAAAATCCCTTAC 230
QY 562 AATATAGCTCTATCAAAAATAAAA 586
Db 229 AACAGGGAAGAACATAATTAATA 205

RESULT 8

PCT-US92-06617A-10/c
; Sequence 10, Application PC/TUS9206617A
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarty, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800

; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, Norval B.
; REGISTRATION NUMBER: 33,595
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-856-7180
; TELEFAX: 312-856-4972
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear

PCT-US92-06617A-10
Query Match 5.28; Score 33.6; DB 5; Length 630;
Best Local Similarity 51.74; Pred. No. 0.44; Mismatches 0; Indels 0; Gaps 0;
Matches 75; Conservative 0;

QY 442 GCGGCTGTATCGGGGTTCTGCTCTTTCCACCTTTTGGTTTATCTCAGTAACATGAC 501
Db 349 GNGTCGATATCGATTACGCTGGTTCTTAATTTTATCTCATCCAGTGAAGATTC 290
QY 502 TCTCTCTTCAGCGTATTATCTTACGACCTTCTAGTTTGTATATCATTCGATGAAGAC 561
Db 289 TTTTCAACATTAGTTTAACTTCTAGTTTATTTCTTATATCTTCCATCAAAATCCCTTAC 230
QY 562 AATATAGCTCTATCAAAAATAAAA 586
Db 229 AACAGGGAAGAACATAATTAATA 205

RESULT 9

US-09-632-098-6/c
; Sequence 6, Application US/09632098
; Patent No. 6420154
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.

; APPLICANT: Baidur, Nand
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
; FILE REFERENCE: 99-39
; CURRENT APPLICATION NUMBER: US/09/632,098
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate sequence
; NAME/KEY: misc_feature
; LOCATION: (1)...(2439)
; OTHER INFORMATION: n = A,T,C or G
US-09-632-098-6

Query Match 5.0%; Score 31.8; DB 4; Length 2439;

Best Local Similarity 27.4%; Pred. No. 3.3; Mismatches 225; Indels 1; Gaps 1;
Matches 104; Conservative 49;

QY 54 TGTCTCTGGATTGGACAAGTATTCTTTCAAAATCAATCTACGCGAGCATGTTCTGTAA 113
Db 2096 TGTTCGTCGTCGACGCGCCGTCATNSMCCNCCRAANCCNGGTTTTCRCAR 2037
QY 114 CACTGGAACGACCAACACCTTCCGCAATTTAGGTAAAGAACTGGTATGGCAACCTTTGT 173
Db 2036 AANGGNGGCCCAACCCGNGCARTGRCARTTGTGTTNSWRTTRCANACNCCRTGN 1977
QY 174 GATTGACTTTTCAAGGAACCCCTAGCAACGCTCTCCGATTATTTTTCATCTACAAGG 233
Db 1976 SWRTGRCANGNGTNRANCAKTYTGNARYCTYTGRAANGCRT-TYTNCKRANCKNCK 1918
QY 234 CGTTTCCTCTCATCTTTTGGACTTTTGGCTGTTATCGGCATACCTTCCCTATCTTGC 293
Db 1917 NSWYTGRCANACCATNCKNGNCCRCAYTNGTNCNGGYTCNACNARNCCNARNCCNAR 1858
QY 294 AGATTTAAAGGTGTAAGGCTGTCCGCAACAGCTGCTGGAGTGATTTTCGGATTTGGGCC 353
Db 1857 NARRTCNARYTGNCNSWNGNARNGCNCCNCKRCANGTNACYTCYTGNCCTC 1798
QY 354 TATCTTCTCTCTACCTTGGATTATCTTCTTTGGAGCTCTCTATCTTGGCAGTATGAT 413
Db 1797 NARRTGNACGNTNSWRTCNACNGNACCATRTGNGNGCNARNARNNSWNGGTTTNCNCC 1738
QY 414 TTCACCTGTCTAGTGTCACA 432
Db 1737 YTGRCAYTGNARYTTNCCR 1719

RESULT 10

US-08-714-402-1/c
; Sequence 1, Application US/08714402
; Patent No. 5910441
; GENERAL INFORMATION:
; APPLICANT: ROCHA, Claudia
; APPLICANT: FISCHETTI, Vincent A.
; TITLE OF INVENTION: FIBRONECTIN AND FIBRINOGEN BINDING
; TITLE OF INVENTION: PROTEIN FROM GROUP A STREPTOCOCCI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS


```
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714.402
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 016921-097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3531 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-714-402-1

Query Match          5.0%; Score 31.8; DB 2; Length 3531;
Best Local Similarity 57.6%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 488 TCAGTAACATGACTCTCTTCATCGCTATATCTTACGACTGCTAGTTGATTATCA 547
DB 3473 TTAGTATCTCGACTTCGGTTTACACCAAGAAATCCCGTACATGCCAGCTTTGGGTATCC 3414

QY 548 TTCGTACATAGCAATAGCTCTATCAAAAATAAA 586
DB 3413 TCTGTGTGACAAACCTGCCCTGACCAACCAATGATA 3375

RESULT 11
US-09-007-005-3/c
; Sequence 3, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 277
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; US-09-007-005-3

Query Match          4.9%; Score 31.6; DB 4; Length 277;
Best Local Similarity 24.0%; Pred. No. 1.2;
Matches 49; Conservative 63; Mismatches 92; Indels 0; Gaps 0;

QY 215 TTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 274
DB 264 TTTTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 274
QY 275 ATACCTTCCCTATCTTTCAGGATTTAAAGGTGCTGAAGCTGCGCAACCAAGTCTGAG 334
DB 204 TCTATCTTCCCTATCTTTCAGGATTTAAAGGTGCTGAAGCTGCGCAACCAAGTCTGAG 334
QY 335 TGAATTTTCGGATTTGCGCCCTATCTTCTCATCTTGGACTTTGGCTGTTATCGGCC 394
DB 144 TTTTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 394
QY 395 TCTATCTTCCCTATCTTTCAGGATTTAAAGGTGCTGAAGCTGCGCAACCAAGTCTGAG 85
DB 84 TTTTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 85

RESULT 12
US-09-244-796-3/c
; Sequence 3, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 277
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; US-09-244-796-3

Query Match          4.9%; Score 31.6; DB 4; Length 277;
Best Local Similarity 24.0%; Pred. No. 1.2;
Matches 49; Conservative 63; Mismatches 92; Indels 0; Gaps 0;

QY 215 TTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 274
DB 264 TTTTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 274
QY 275 ATACCTTCCCTATCTTTCAGGATTTAAAGGTGCTGAAGCTGCGCAACCAAGTCTGAG 334
DB 204 TCTATCTTCCCTATCTTTCAGGATTTAAAGGTGCTGAAGCTGCGCAACCAAGTCTGAG 334
QY 335 TGAATTTTCGGATTTGCGCCCTATCTTCTCATCTTGGACTTTGGCTGTTATCGGCC 394
DB 144 TTTTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 394
QY 395 TCTATCTTCCCTATCTTTCAGGATTTAAAGGTGCTGAAGCTGCGCAACCAAGTCTGAG 85
DB 84 TTTTATTTTTCATCAAGCGTTCTCCTCATCTTGGACTTTGGCTGTTATCGGCC 85

RESULT 13
US-09-068-140A-14/c
; Sequence 14, Application US/09068140A
; Patent No. 6281409
; GENERAL INFORMATION:
; APPLICANT: Mary Rose Woodhead, Mark Andrew Taylor
; APPLICANT: and Rex Michael Brennan
; TITLE OF INVENTION: Blackcurrant Promoters and Genes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
```


COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/068,140A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/04807
FILING DATE: No. 6281409ember 4, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Dinner, Dara L.
REGISTRATION NUMBER: 33,680
REFERENCE/DOCKET NUMBER: C70237
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5017
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5150 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Ribes nigrum
STRAIN: Ben Alder
US-09-068-140A-14

Query Match 4.9%; Score 31.6; DB 4; Length 5150;
Best Local Similarity 50.7%; Pred. No. 5.6; Mismatches 74; Indels 0; Gaps 0;
Matches 76; Conservative 0;
Qy 336 GATTTCGGATTGGCGCTATCTCTCTACCTTCGCGATTATCTTTTGGAGCTCT 395
Db 167 GATTGCCAAATGCTGCTACTTTAGTTAGATCATTAAGTTTCTCTAGTTTCT 108
Qy 396 CTATCTGGCAGTATGATTCTCTCTAGTGTGTCACAGCATCGATTGGCGGTGTTATCGG 455
Db 107 CACAGTTCTTAGCTCTATTCAGTTTATCAGTTGATCGTCCATAGCATGATTCCTCA 48
Qy 456 GTTCTGCTCTTCCACTTTTGGTTTAT 485
Db 47 GTTTTGATGATTCACCGCATGATTTAT 18

RESULT 14
US-09-798-096-10
Sequence 10, Application US/09798096
Patent No. 6399378
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF RECOL2 EXPRESSION
FILE REFERENCE: RTS-0207
CURRENT APPLICATION NUMBER: US/09/798,096
CURRENT FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 89
SEQ ID NO 10
LENGTH: 99500
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-798-096-10

Query Match 4.9%; Score 31.2; DB 4; Length 99500;

Best Local Similarity 46.4%; Pred. No. 35;
Matches 102; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
Qy 289 TTTCAGGATTAAAGGTGGTAAGGCTGTCGCAACACAGTCTCGAGTGTATTTTCGGATTT 348
Db 57656 TGTTCAGCTGTTTCAGTGTGTTGGTGTGTTTTCGAATAATGCTGCTGTGAACATTTGTGA 57715
Qy 349 GCGCTATCTCTGCTCTACCTTGCGATTATCTTCTTTGGAGCTCTCTATCTTTGGCAGT 408
Db 57716 GAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57775
Qy 409 ATGATTTCAGTCTAGTGTGTCACAGCATCGATTGCGGCTGTTATCGGGGTTCTGCTCTTT 468
Db 57776 ATGATTAACTTTTCAATTCAAGGAGTAACTAACTAGTTTTCAAAGTGCCTCGCTCTTTT 57835
Qy 469 CCACCTTTTGGTGTATCTCTGAGTAACTATGACTCTCTCT 508
Db 57836 TTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57875

RESULT 15
US-09-676-610B-24/C
Sequence 24, Application US/09676610B
Patent No. 644465
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Jacqueline Wyatt
APPLICANT: Susan M. Freier
TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
FILE REFERENCE: RTS-0138
CURRENT APPLICATION NUMBER: US/09/676,610B
CURRENT FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 182
SEQ ID NO 24
LENGTH: 169998
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: (1208)...(1472)
NAME/KEY: intron
LOCATION: (1473)...(124390)
NAME/KEY: exon
LOCATION: (124391)...(124544)
NAME/KEY: intron
LOCATION: (124545)...(125409)
NAME/KEY: exon
LOCATION: (125410)...(125595)
NAME/KEY: intron
LOCATION: (125596)...(128711)
NAME/KEY: exon
LOCATION: (128712)...(128848)
NAME/KEY: intron
LOCATION: (128849)...(133400)
NAME/KEY: exon
LOCATION: (133401)...(133469)
NAME/KEY: intron
LOCATION: (133470)...(134652)
NAME/KEY: exon
LOCATION: (134653)...(134773)
NAME/KEY: intron
LOCATION: (134774)...(136116)
NAME/KEY: exon
LOCATION: (136117)...(136261)
NAME/KEY: intron
LOCATION: (136262)...(137936)
NAME/KEY: exon
LOCATION: (137937)...(138053)
NAME/KEY: intron
LOCATION: (138054)...(138637)
NAME/KEY: exon
LOCATION: (138638)...(138766)
NAME/KEY: intron

LOCATION: (138767)...(138864)
NAME/KEY: exon
LOCATION: (138865)...(138940)
NAME/KEY: intron
LOCATION: (138941)...(139765)
NAME/KEY: exon
LOCATION: (139766)...(139860)
NAME/KEY: intron
LOCATION: (139861)...(142245)
NAME/KEY: exon
LOCATION: (142246)...(142445)
NAME/KEY: intron
LOCATION: (142446)...(143605)
NAME/KEY: exon
LOCATION: (143606)...(143738)
NAME/KEY: intron
LOCATION: (143739)...(145838)
NAME/KEY: exon
LOCATION: (145839)...(145931)
NAME/KEY: intron
LOCATION: (145932)...(147385)
NAME/KEY: exon
LOCATION: (147386)...(147544)
NAME/KEY: intron
LOCATION: (147545)...(153274)
NAME/KEY: exon
LOCATION: (153275)...(153321)
NAME/KEY: intron
LOCATION: (153322)...(155088)
NAME/KEY: exon
LOCATION: (155089)...(155231)
NAME/KEY: intron
LOCATION: (155232)...(156025)
NAME/KEY: exon
LOCATION: (156026)...(156151)
NAME/KEY: intron
LOCATION: (156152)...(156826)
NAME/KEY: exon
LOCATION: (156827)...(156928)
NAME/KEY: intron
LOCATION: (156929)...(163399)
NAME/KEY: exon
LOCATION: (163400)...(163586)
US-09-676-610B-24

Query Match 4.98; Score 31.2; DB 4; Length 16998;
Best Local Similarity 46.18; Pred. No. 46;
Matches 105; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 413 TTTCACGTCTAGTGTACAGCATCGATTGGGCTGTTATCGGGGTCTCTCTTTCCAC 472
Db 167319 TCTCATGGTCTACTGGTCTCATATGGATTCCTCCAGACACCTCTGTTTTTTCCTTT 167260
QY 473 TTTTGGTTTATCTCAGTAACATATGACCTCTCTCTTCATCGCTATTATCTTAGCACTTG 532
Db 167259 TTCTTTTGTCTAGTTGATATATTTTATGTAATCTCTTTGGCAGTTTAGTCAGGACAAT 167200
QY 533 CTAGTTTGATATCATTCGTCATAAGGACAATATAGCTGTATCAAAAATAAACTGAAA 592
Db 167199 GAAACTTGCATAACAGTGGCCAGCTCTCCCATCAACATTTAGAAGAAACTTACCTAGAT 167140
QY 593 ATTGTGTCCTTGGGATTGCAACTAACCCATCAAGATCCTAAAAAT 640
Db 167139 AGTTTGTGCTTTGTGAGTCAACAGGACATTTATAAGCCATAAAT 167092

Search completed: December 25, 2002, 14:24:38
Job time : 177.625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 ; Search time 189.61 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues

Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications, NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	642	100.0	642	12	US-10-068-080-2
2	640.4	99.8	642	10	US-09-815-242-9231
3	640.4	99.8	642	10	US-09-815-242-9231
c 4	193.2	30.1	6691	10	US-09-070-927A-88
5	102.4	16.0	609	10	US-09-823-246-1
6	100	15.6	582	12	US-10-068-080-4
7	99.2	15.5	605	10	US-09-815-242-4250
8	99.2	15.5	609	10	US-09-815-242-8039
9	67.4	10.5	600	10	US-09-815-242-6900
10	59	9.2	570	10	US-09-815-242-7702
11	57.8	9.0	612	10	US-09-815-242-9659
c 12	56.6	8.8	189	10	US-09-815-242-3464
13	53.6	8.3	618	10	US-09-815-242-6206
14	38.4	6.0	618	10	US-09-815-242-7616
c 15	36.6	5.7	1118	9	US-09-938-842A-4649
16	36.4	5.7	358	10	US-09-783-590-5896
17	34.4	5.4	400	10	US-09-960-352-2619
18	34.4	5.4	403	10	US-09-960-352-4074
19	34.4	5.4	424	10	US-09-960-352-11631

20	34.4	5.4	430	10	US-09-960-352-9747	Sequence 9747, Ap
c 21	34	5.3	659158	9	US-09-771-208-20	Sequence 20, Appl
22	33.2	5.2	326014	10	US-09-731-231A-3	Sequence 3, Appl
23	32.6	5.1	408	10	US-09-960-352-1221	Sequence 1221, Ap
c 24	32.4	5.0	5621	10	US-09-070-927A-571	Sequence 571, App
c 25	32.4	5.0	6855	10	US-09-764-864-1694	Sequence 1694, Ap
26	32.4	5.0	21045	10	US-09-764-864-1695	Sequence 1695, Ap
27	32.2	5.0	204	9	US-10-040-739-505	Sequence 505, App
c 28	32.2	5.0	3087	10	US-09-815-242-4830	Sequence 4830, Ap
c 29	32.2	5.0	3144	10	US-09-815-242-8985	Sequence 8985, Ap
c 30	31.8	5.0	9439	10	US-09-764-877-2224	Sequence 2224, Ap
c 31	31.6	4.9	718	9	US-10-001-857-107	Sequence 107, App
c 32	31.6	4.9	988	9	US-10-016-157A-29	Sequence 29, Appl
c 33	31.6	4.9	1671	10	US-09-815-242-5959	Sequence 5959, Ap
c 34	31.6	4.9	5150	10	US-09-800-528-14	Sequence 14, Appl
35	31.4	4.9	471	9	US-09-946-807-1424	Sequence 1424, Ap
36	31.4	4.9	471	10	US-09-795-668-1424	Sequence 1424, Ap
37	31.4	4.9	471	10	US-09-795-668-1424	Sequence 1424, Ap
38	31.4	4.9	598	10	US-09-864-761-12491	Sequence 12491, A
39	31.4	4.9	17493	10	US-09-804-471A-3	Sequence 3, Appl
40	31.2	4.9	166	10	US-09-770-696-376	Sequence 376, App
41	31.2	4.9	264	10	US-09-969-373-472	Sequence 472, App
c 42	31.2	4.9	597	9	US-09-938-842A-766	Sequence 766, App
43	31.2	4.9	717	10	US-09-770-149-172	Sequence 172, App
c 44	31.2	4.9	1416	10	US-09-925-299-173	Sequence 173, App
c 45	31.2	4.9	197496	9	US-09-877-177-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-10-068-080-2
; Sequence 2, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritze, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(639)
US-10-068-080-2

Query Match	100.0%	Score	642	DB	12	Length	642
Best Local Similarity	100.0%	Pred. No.	3.3e-180				
Matches	642	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	ATGATTACAATAGTTTATTAACTAGCTATCTGCTGGTTCGATTCCTGCTGCTC	60				
Db	1	ATGATTACAATAGTTTATTAACTAGCTATCTGCTGGTTCGATTCCTGCTGCTC	60				
Qy	61	TGGATTGGACAAGTATCTTCAATCAATCTACGCGAGCATGTTCTGTTAAGTACGGA	120				
Db	61	TGGATTGGACAAGTATCTTCAATCAATCTACGCGAGCATGTTCTGTTAAGTACGGA	120				
Qy	121	ACGACCAACACCTTCCGCACTTTTAGTATAGGTAAGAAAGCTGGTATGGAACCTTTGATTGAC	180				
Db	121	ACGACCAACACCTTCCGCACTTTTAGTATAGGTAAGAAAGCTGGTATGGAACCTTTGATTGAC	180				
Qy	181	TTTTTCAAGGAACCTTCCGCACTTTTAGTATAGGTAAGAAAGCTGGTATGGAACCTTTGATTGAC	240				
Db	181	TTTTTCAAGGAACCTTCCGCACTTTTAGTATAGGTAAGAAAGCTGGTATGGAACCTTTGATTGAC	240				

Db 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTATTTTCATCTACAGGCGTTTCT 240
Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 300
Db 241 CCTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 300
Qy 301 AAAGTGGTAAAGCTGTCGCAACAGTGGAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 360
Db 301 AAAGTGGTAAAGCTGTCGCAACAGTGGAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 360
Qy 361 TGTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 420
Db 361 TGTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 420
Qy 421 TCTAGTGTACAGCATCGATTGGCGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 480
Db 421 TCTAGTGTACAGCATCGATTGGCGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 480
Qy 481 TTTATCTCTGAGTAAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 540
Db 481 TTTATCTCTGAGTAAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 540
Qy 541 ATTATCATCTTCTATAGGACAATATAGCTCGTATCAAAATAAACTGAAAAATTTTGGTC 600
Db 541 ATTATCATCTTCTATAGGACAATATAGCTCGTATCAAAATAAACTGAAAAATTTTGGTC 600
Qy 601 CCTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAATAA 642
Db 601 CCTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAATAA 642

RESULT 2

US-09-815-242-9231
; Sequence 9231, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9231
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(642)
US-09-815-242-9231

Query Match 99.8%; Score 640.4; DB 10; Length 642;
Best Local Similarity 99.8%; Pred. No. 9.8e-180;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGATTACAATAGTATTTTATTAACTTCTAGCTATCTGCTGGTTCGATTCCATCTCGGTCTC 60
Db 1 ATGATTACAATAGTATTTTATTAACTTCTAGCTATCTGCTGGTTCGATTCCATCTCGGTCTC 60
Qy 61 TGAATTGGACAAGTATTTCTTCAAACTACGAGCATGTTCTTGGTAACTGGA 120
Db 61 TGAATTGGACAAGTATTTCTTCAAACTACGAGCATGTTCTTGGTAACTGGA 120
Qy 121 ACAGCAACACCTTCCCATTTTAGTAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 180
Db 121 ACAGCAACACCTTCCCATTTTAGTAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 180
Qy 181 TTTTCAAGGAACCTTAGCAAGCTGCTTCCGATTTATTTTTCATCTACAGGCGTTTCT 240
Db 181 TTTTCAAGGAACCTTAGCAAGCTGCTTCCGATTTATTTTTCATCTACAGGCGTTTCT 240
Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 300
Db 241 CCTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 300
Qy 301 AAAGTGGTAAAGCTGTCGCAACAGTGGAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 360
Db 301 AAAGTGGTAAAGCTGTCGCAACAGTGGAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 360
Qy 361 TGTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 420
Db 361 TGTCTCATCTTTGGACTTTTGGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 420
Qy 421 TCTAGTGTACAGCATCGATTGGCGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 480
Db 421 TCTAGTGTACAGCATCGATTGGCGCTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 480
Qy 481 TTTATCTCTGAGTAAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 540
Db 481 TTTATCTCTGAGTAAGTGTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 540
Qy 541 ATTATCATCTTCTATAGGACAATATAGCTCGTATCAAAATAAACTGAAAAATTTTGGTC 600
Db 541 ATTATCATCTTCTATAGGACAATATAGCTCGTATCAAAATAAACTGAAAAATTTTGGTC 600
Qy 601 CCTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAATAA 642
Db 601 CCTTGGGATTGAACCTTAACCCATCAAGATCCTAAAAATAA 642

RESULT 3

US-09-815-242-9438
; Sequence 9438, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 9438
;; LENGTH: 642
;; TYPE: DNA
;; ORGANISM: Streptococcus pneumoniae
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(642)
US-09-815-242-9438

Query Match 99.8%; Score 640.4; DB 10; Length 642;
Best Local Similarity 99.8%; Pred. No. 9.8e-180;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGATTACAAGTATTTTAAATCTAGCCTATCTGCTGGGTTCGATTCCTCCATCTGGTCTC 60
Db 1 ATGATTACAAGTATTTTAAATCTAGCCTATCTGCTGGGTTCGATTCCTCCATCTGGTCTC 60

Qy 61 TGGATTGGCAAGATATCTTTCAATCAATCTACCGGAGCATGGTTCGTGTAACACTGGA 120
Db 61 TGGATTGGCAAGATATCTTTCAATCAATCTACCGGAGCATGGTTCGTGTAACACTGGA 120

Qy 121 ACGACCAACACTTCCCGCATTTAGTAAAGAGCTGGTATGGCAACCTTTGTGATTCAC 180
Db 121 ACGACCAACACTTCCCGCATTTAGTAAAGAGCTGGTATGGCAACCTTTGTGATTCAC 180

Qy 181 TTTTCAAGGAACCTTAGCAAGCTGCTCCGATTTATTTTCATCTACAAGCGCTTCT 240
Db 181 TTTTCAAGGAACCTTAGCAAGCTGCTCCGATTTATTTTCATCTACAAGCGCTTCT 240

Qy 241 CCTCTCATCTTTGGACTTTTGGCTTTATCGGCCATACCTTCCCTATCTTTGCAGGATTT 300
Db 241 CCTCTCATCTTTGGACTTTTGGCTTTATCGGCCATACCTTCCCTATCTTTGCAGGATTT 300

Qy 301 AAAGTGTAGGCTGTCCCAACCACTGCTGGAGTGATTTCCGATTTCCGCTATCTTC 360
Db 301 AAAGTGTAGGCTGTCCCAACCACTGCTGGAGTGATTTCCGATTTCCGCTATCTTC 360

Qy 361 TGCTCTACCTTTGGCATTTCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
Db 361 TGCTCTACCTTTGGCATTTCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420

Qy 421 TCTAGTGTACAGCATGATTCGGCTGTTATCGGGGTTCTGCTCTTTCCACTTTTGGT 480
Db 421 TCTAGTGTACAGCATGATTCGGCTGTTATCGGGGTTCTGCTCTTTCCACTTTTGGT 480

Qy 481 TTTATCTTGAGTAAGTATCTCTCTCATCGCTATTTCTTAGCACTTTGCTAGTTTG 540
Db 481 TTTATCTTGAGTAAGTATCTCTCTCATCGCTATTTCTTAGCACTTTGCTAGTTTG 540

Qy 541 ATTATCATCTGTCATAGGACAATATAGCTGTATCAAAATATAAATGAAATTTTGCTC 600
Db 541 ATTATCATCTGTCATAGGACAATATAGCTGTATCAAAATATAAATGAAATTTTGCTC 600

Qy 601 CTTGGGATTTGAACCTAACCCATCAAGATCTTAAAAATAA 642
Db 601 CTTGGGATTTGAACCTAACCCATCAAGATCTTAAAAATAA 642

RESULT 4
US-09-070-927A-88/c
; Sequence 88, Application US/09070927A
; Patent No. US20020120116A1

;; GENERAL INFORMATION:
;; APPLICANT: Charles A. Kunsch
;; Patrick J. Dillon
;; Steven Barash
;; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptid
;; NUMBER OF SEQUENCES: 982
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Human Genome Sciences, Inc.
;; STREET: 9410 Key West Avenue
;; CITY: Rockville
;; STATE: Maryland
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII Text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/070,927A
;; FILING DATE: 04-May-2000
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/046,655
;; FILING DATE: 1997-05-16
;; APPLICATION NUMBER: 60/044,031
;; FILING DATE: 1997-05-06
;; APPLICATION NUMBER: 60/066,009
;; FILING DATE: 1997-11-14
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kenley K. Hoover
;; REGISTRATION NUMBER: 40,302
;; REFERENCE/DOCKET NUMBER: PB369
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 88:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6691 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-070-927A-88

Query Match 30.1%; Score 193.2; DB 10; Length 6691;
Best Local Similarity 58.3%; Pred. No. 4.6e-47;
Matches 358; Conservative 0; Mismatches 253; Indels 3; Gaps 1;

Qy 1 ATGATTACAAGTATTTTAAATCTAGCCTATCTGCTGGGTTCGATTCCTCCATCTGGTCTC 60
Db 5424 ATGAAAATCGTCATTTTGTCTTACTTTGCTCCCTATTTATTAGGTTCGATTCCTCCATCTGGTCTC 5365

Qy 61 TGGATTGGCAAGATATCTTTCAATCAATCTACCGGAGCATGGTTCGTGTAACACTGGA 120
Db 5364 TGGATTGGTAACATCTTTCTTTAAAGAGATATACGCCAATTTGGAGTGGGAATACAGGA 5305

Qy 121 ACGACCAACACTTCCCGCATTTTAGTAAAGAGCTGTATGGCAACCTTTGTGATTCAC 180
Db 5304 ACAACCAATACATTTCTGCTCTAGGGAACCTTCGCGAATACGGTATTTATATGAT 5245

Qy 181 TTTTCAAGGAACCTTAGCAAGCTGCTTCCGATTTATTTTCATCTACAAGGGCTTTCT 240
Db 5244 ATCTTGAAGGAACCTTAGCAAGCTTATACCCCTATTTGTTGGTTTACAAGGGCTGAAT 5185

Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
Db 5184 CCGCTCTTTTGGGGTAGCAGCTGTTTGGGGCATCTTTTCCCTATTTTGGCCAAATTTTC 5125

Qy 301 AAAGTGTAGGCTGTCCCAACCACTGCTGGAGTGATTTTTCGATTTTTCGGCTATCTTC 360
Db 5124 AAAGTGTAGGCTGTCCCAACCACTAGCCTGCGCATGTTATTAGCATACAGCCCACTTT 5065

Qy	361	TGTCCTACCTCGGATTAATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG	420
Db	5064	TTTATTATCTGCTCTATTTTGTGTGATTTGCTGTGATCTAAACAGCATGGTGAGTTTA	5005
Qy	421	TCFAGTGTACACAGCATCGATTGGGCGTGTATCGGGGTCTGCTCTTTCCACT---	477
Db	5004	ACAAGTATGATTAGCGCTGTACTAATTACACTTTTCTACTATATTTGCCCTTTCACGTGC	4945
Qy	478	GGTTTATCTGAGTAACATATGACTCTCTCTTCATCGCTATATTCTTAGCACTTGCCTAGT	537
Db	4944	CCAGCTATTTTACCAACATTTAACTGGTTATTAAACCGTGATTCGGGATTCGCTTGCCTACT	4885
Qy	538	TTGATTATCATCTGCATAGGACAAATATAGCTCGCTATCAAAAATAAAACCTGAAAATTTG	597
Db	4884	TTTATCTTTCCGCTCATCGTGAACAACATTCAGCGAATTTAAATATGCACAGAAAGTCGT	4825
Qy	598	GTCCCTTGGGATT	611
Db	4824	CTTCTTTTGGCTT	4811

```

RESULT 5
US-09-823-246-1
> Sequence 1, Application US/09823246
> Patent No. US20020058789A1
> GENERAL INFORMATION:
> APPLICANT: Burnham, Martin K. R.
> APPLICANT: Biswas, Sanjoy
> APPLICANT: Zalacain, Magdalena
> APPLICANT: Warren, Patrick V.
> APPLICANT: Slyvester, Daniel R.
> APPLICANT: McDevitt, Damien
> TITLE OF INVENTION: vnes
> FILE REFERENCE: GM20001
> CURRENT APPLICATION NUMBER: US/09/823,246
> CURRENT FILING DATE: 2001-03-30
> PRIOR APPLICATION NUMBER: US 60/193,496
> PRIOR FILING DATE: 2000-03-31
> NUMBER OF SEQ ID NOS: 2
> SOFTWARE: FastSeq for Windows Version 4.0
> SEQ ID NO 1
> LENGTH: 609
> TYPE: DNA
> ORGANISM: Staphylococcus aureus
US-09-823-246-1

```

Query Match	16.0%	Score 102.4;	DB 10;	Length 609;
Best Local Similarity	52.6%;	Pred. No. 9.9e-21;		
Matches 266;	Conservative	0; Mismatches 216;	Indels 24;	Gaps 1;

Qy	1	ATGATTACAAGTATTTTAATACTCCTACCCATCTGCTGGGTTCGATTCCATCTGGTCTC	60
Db	1	ATGATGAATACGTCATGTTACTAAGTAGTATCTTATCGGCGCTTCCCAGGTGGATTC	60
Qy	61	TGGATTGGACAAAGTATCTTTCAATCAATCTACGCCAGCATGGTTCCTGGTAAACATCGGA	120
Db	61	GTAATTGC AAAAATATTTTTCAA AAAAGATATTAGACAATTTGGTAGTGSTTA TACTGCG	120
Qy	121	AGGACCACACGCTCCCGCATTTTAGTGAAGAAGCTGGTATGGGC AACCTTTGTGATGAC	180
Db	121	GCTACTAATAGCTTTAGTAGTATTAGTGCCTCGCAGGATCTTGGTAACATTTCTTAGAT	180
Qy	181	TTTTTCAAAGAACCCCTAGCAAGCGTCTCCGATTA TTTTTCATCTACAAGGCGTTTCT	240
Db	181	ATTTTCAAAGGGTCA TAAC TGTTTCTTCCCTTTATGGTTACAAGTTTCACG CAGATGGC	240
Qy	241	CCTCTC-----ATCTTTTGAC TTTTGGCTGT TTATCGGCCAT	276
Db	241	CCTATTAGTACTTTTTTACAAA TGGTTTAAT GTTGGCTTATTCGCTAT ACTTGGACAC	300
Qy	277	ACCTCCCTATCTTTTGCAGAGATTTAAAGGTGGTGAAGCGT GTCGGCAACCAAGTCTGGAGTG	336
Db	301	GTTTATCCCTGTTATTTAAAAATCCAAGGTGCGAAGCTGTTC AACTAGTGCAGGTGCTC	360

QY 337 ATTTTGGAGTTGGCCGCTATCTTCGTCTCTACCTTGCGATTATCTCTTTTGAGCTCTC 396
 ||| |||| | | | | | | | | | | | | |
 DB 361 GCTTTGGGAGTCAATCCGATACTTTTAACTACTTGC AATATATCTCTTTATTGTATTG 420
 QY 397 TATCTTGGCAGTAGTATGATTTTCATCTGCTAGTGTGCACAGCATCGATTCGGGCTGTTATCGGG 456
 || | | | | | | | | | | | | | | | | |
 DB 421 AGAATTTTTTAATAATATGTTTCTTTAGCAAGTATCGTTGCGACAATTTGCTGTGTGATTGGC 480
 QY 457 GTCCTGCTCTTTTCCACTTTTTCGGTTT 482
 || | | | | | | | | | | | | | | | | |
 DB 481 TCGCTTATCATTTCAAGACTATATTTT 506

RESULT 6
US-10-068-080-4
; Sequence 4, Application US/10068080
; Patent No.. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(579)
US-10-068-080-4

Query Match	15.6%;	Score 100;	DB 12;	Length 582;
Best Local Similarity	52.9%;	Pred. No. 4.9e-20;		
Matches 239;	Conservative 0;	Mismatches 210;	Indels 3;	Gaps 1;

QY	1	ATGATTACAATAGTTTTTATTATCTACGCTATCTGCTGGGTTCGATTCATCTCGTCTC 60
DB	1	ATGTTAAATGCTTTATTTGATTATTTTGGCTACTTTGATAGGCAGCATTTCCATCTGGCTTA 60
QY	61	TGGATTGGCAGAGTATCTTTCAAAATCAATCTACCGCAGCATGGTTCGGTAACACTGGA 120
DB	61	ATTGTGGGCAAGCTTGGCCAAAGGAATGATATTCGGGAGCAGGAAGCGGCAAACTTAGCG 120
QY	121	ACGACCAACACCTCCCGCATTTTAGTGAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 180
DB	121	GCTACCAATGCATTCGGTACATGGGTGTAAAGGCTGGTTCGGTCGTATAGCCGGAGAT 180
QY	181	TTTTTCAAGGAACCTTAGCAACGCTGCTTCGGATATATTTTTCATCTACAAGCGCTTTCT 240
DB	181	ATTTTGAAGGGACACTGGCAACTGCATTTGCCCTTTTCTCATGCATGT---TGAATATTCAC 237
QY	241	CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTCAGCAGATTT 300
DB	238	CCGCTTCTTTCGACAGAGTCTTTTCGGGTTTTCGCGGCTTTAGGCCACAGTGTTCCTCCCATCTTCGCCAAATTT 297
QY	301	AAAGGTGGTAAAGGCTTCGCAACACAGTGTGGAGTGATTTTCGGGATTTTCGGCTATCTTTC 360
DB	298	AAAGCGGTAAAGCGCTGGCGACATCAGAGCGGTTTGGCTATTTTTCGACACCCCTGTTA 357
QY	361	TGCTCTCTACCTTTCGGAATATCTTCTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 420
DB	358	TTTATCAGCATGGTTTCGGGTATCTTCTCATCTTTTTTATCTACTGACATAAATTCGTTTCTCTC 417
QY	421	TCTAGTGTACAGCATTCGATTCGGCGCTGTTAT 452
DB	418	TCATTCGATGTTTAAACGGGATCTATACCTGTTAT 449

RESULT 7

US-09-815-242-4250
; Sequence 4250, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4250

; LENGTH: 606

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-4250

Query Match 15.5%; Score 99.2; DB 10; Length 606;

Best Local Similarity 52.2%; Pred. No. 8.7e-20;

Matches 264; Conservative 0; Mismatches 218; Indels 24; Gaps 1;

QY 1 ATGATTACAAATAGTTTATTAATCTAGCTATCTGCTGGGTTGCGATTCATCTGCTC 60

DB 1 ATGATGATAATCGTCATGTTACTACTAAGTTATCTATCGGCGCTTCCCAAGTGGATTC 60

QY 61 TGGATTGGCAAGTATTCTTTCAATCAATCTACCGGAGCATGGTCTGGTAACACTGGA 120

DB 61 GTAATTTGGAATAATTTTTCAAAAGAGATATTAGACAATTTGGTAGTGGTAATACTGGC 120

QY 121 ACGACCAACACCTTCGCGATTTTAGTGAAGAGCTGGTATGCGAACCTTTCTGATTGAC 180

DB 121 GCTACTAATAGCTTTAGAGTATTAGTCTCTCGAGGATTTCTTGGTAACATTTCTAGAT 180

QY 181 TTTTCAAGAGCAACCTAGCAACGCTGCTTCGCAATTTATTTTCAATCAACAGCGTTTCT 240

DB 181 ATTTTCAAGAGGTTTCATAACTGTTTCTTCCCTTTATGTTTACCAGTTCCACGATGCG 240

QY 241 CCTCTC-----ATCTTTGGACTTTTGGCTGTTATCGGCCAT 276

DB 241 CTTATTAGTACTTTTTTACAAATGGTTTAAATTTGGTCTTATTCGCTACTTGGACAC 300

QY 277 ACTTCCCTATCTTTTCAGGATTTAAAGTGGTGAAGGCTGTCGCAACGAGTCTGAGTG 336

DB 301 GTTTATCTGTTTATTTAAATTCAGGTGGCAAGCGGTGCAACTAGTACGAGTGC 360

QY 337 ATTTTCGGATTTGCGCTATCTTCTCTACTACCTTCGCAATTTCTTTTGGAGCTCTC 396

DB 361 GTCTTGGAGTCAACCCGACTACTTTTACTAATACTTTGCAATTTACTCTTTATTGATTG 420

QY 397 TATCTTGGCAGTATGATTTTCACTGTCTAGTGTACAGCATGATTTGGCTGGTGTATCGG 456

DB 421 AAGATTTTAAATATGTTTCTTACCAAGTATCTGTTGAGCAATTTGCTGTGATGGC 480

QY 457 GTTCTGCTCTTTTCCACATTTTGGTTT 482

DB 481 TCGCTTATCATTCATTCAGAGCTATATTTT 506

RESULT 8

US-09-815-242-8039

; Sequence 8039, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8039

; LENGTH: 609

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(609)

US-09-815-242-8039

Query Match 15.5%; Score 99.2; DB 10; Length 609;

Best Local Similarity 52.2%; Pred. No. 8.7e-20;

Matches 264; Conservative 0; Mismatches 218; Indels 24; Gaps 1;

QY 1 ATGATTACAAATAGTTTATTAATCTAGCTATCTGCTGGGTTGCGATTCATCTGCTC 60

DB 1 ATGATGATAATCGTCATGTTACTACTAAGTTATCTATCGGCGCTTCCCAAGTGGATTC 60

QY 61 TGGATTGGCAAGTATTCTTTCAATCAATCTACCGGAGCATGGTCTGGTAACACTGGA 120

DB 61 GTAATTTGGAATAATTTTTCAAAAGAGATATTAGACAATTTGGTAGTGGTAATACTGGC 120

QY 121 ACGACCAACACCTTCGCGATTTTAGTGAAGAGCTGGTATGCGAACCTTTCTGATTGAC 180

DB 121 GCTACTAATAGCTTTAGAGTATTAGTCTCTCGAGGATTTCTTGGTAACATTTCTAGAT 180

QY 181 TTTTCAAGAGCAACCTAGCAACGCTGCTTCGCAATTTATTTTCAATCAACAGCGTTTCT 240

DB 181 ATTTTCAAGAGGTTTCATAACTGTTTCTTCCCTTTATGTTTACCAGTTCCACGATGCG 240

Qy	241	CCTCTC-----ATCTTTGGACTTTTGGCTGTATTGCGCCAT	276
Dd	241	CCTATTAGTACTTTTTACAAATGGTTTTAAITTTGGCTTATTTCCTACTATTGGACAC	300
Qy	277	ACCTTCCTCATCTTTGCAGGATTTAAAGTGTAAGSCTGTCCGAACCACTGCTGGAGTG	336
Dd	301	GTTTATCCTGTTATTAAATTCRAAGTGGCAAGCGTTGCCAATACTAGTCGAGTGTC	360
Qy	337	ATTTTCGGATTTGGGCTATCTCTCTGTCCTACCTTCGCAATATCTTCTTTGGAGCTCTC	396
Dd	361	GTCTTGGGAGTCAACC CGAT ACTTTACTAATACTTAGCAATATCTTCTTTATTGTATTG	420
Qy	397	TATCTTGGCAGTAGTTTACCTGCTAGTGTCACAGCATCGAATTCGGGCTGTTATCGGG	456
Dd	421	AAGATTTTAAATATGTTCTTTAGCAAGTATCGTTGCACAATTTGCTGTGTGATTGGC	480
Qy	457	GTCTGCTCTTTCCACTTTTGGTTT	482
Dd	481	TCGCTTATCATTTCAAGACTATATTTT	506

RESULT 9

RES-002-153
US-500-815-242-6900
; Sequence 6900, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert.
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Travick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

1 TITLE OF INVENTION: Identification of Essential Genes in
2 TITLE OF INVENTION: Prokaryotes
3 FILE REFERENCE: ELITRA.011a
4 CURRENT APPLICATION NUMBER: US/09/815,242
5 CURRENT FILING DATE: 2001-03-21
6 PRIOR APPLICATION NUMBER: 60/191,078
7 PRIOR FILING DATE: 2000-03-21
8 PRIOR APPLICATION NUMBER: 60/206,848
9 PRIOR FILING DATE: 2000-05-23
10 PRIOR APPLICATION NUMBER: 60/207,727
11 PRIOR FILING DATE: 2000-05-26
12 PRIOR APPLICATION NUMBER: 60/242,578
13 PRIOR FILING DATE: 2000-10-23
14 PRIOR APPLICATION NUMBER: 60/253,625
15 PRIOR FILING DATE: 2000-11-27
16 PRIOR APPLICATION NUMBER: 60/257,931
17 PRIOR FILING DATE: 2000-12-22
18 PRIOR APPLICATION NUMBER: 60/269,308
19 PRIOR FILING DATE: 2001-02-16
20 NUMBER OF SEQ ID NOS: 14110
21 SOFTWARE: FASTSEQ for Windows Version 4.0

```

Query Match 10.5%; Score 67.4; DB 10; Length 600;
Best Local Similarity 50.5%; Pred. No. 2.2e-10;
Matches 164; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

[illegible]

Db	80	TTGGAGGCTTCCAGATCCTCGTCAAAATGGTTACACACACCTGGGGGACCAATGTCT	139
Qy	134	TCCGCATTTTAGTAAAGACGTGGTATGGCAACCTTTTGATGACTTTTTCACAGGAA	193
Db	140	TCCGCATTTGGAAATCGTAAGCTCGCCTTGGCGGTCTTATTTTGTATGTTTAAAGGGA	199
Qy	194	CCCTTAGCAACGCTGCTTCCGATTATTTTTCATCTACAAGCGTTCCTCTCTCATCTTGG	253
Db	200	TCATTCTGTATGGCAGGGTATATTATTAGTGTTTAACCCAATTTGAATGGGAATG	259
Qy	254	GACTTTTGGCTGTATCGGCCATACCTTCCCTCATCTTGCAGGATTTAAAGTGGTAAGG	313
Db	260	PCCTTGGTGTGTTTAGGCGCATATTTTCCCAATTTTCTTTCATTTAAAGGTGGCAAG	319
Qy	314	CTGTCCCAACGAGTCTGGAGTGAT	338
Db	320	GTGTTGCAACTGCCCTTTGGGGCAAT	344

RESULT 10

```

US-09-815-242-7702
: Sequence 7702, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Ess
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7702
: LENGTH: 570
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(570)
: US-09-815-242-7702

```

Query Match	9.28;	Score 59;	DB 10;	Length 570;
Best Local Similarity	48.98;	Pred. No. 6.4e-08;		

Qy	22	ATCTAGCCTATTCGCTGGGTCGATTCCATCTGGTCTCTGGATTGGACAAAGTATTCTTT	81
Db	19	ATCCTCGCCTACCTGCTCGGTTCCCTCTCCTTCGCCGTCTGCTCAGCCGCTGTTCCGC	78
Qy	82	CAAAATCAATCTACCGGAGCATGGTTCTGGTAACACTGGAAGACCAACACCTTCCGCAT	147

Db 79 ACCAGGACCGCGCGGCTTCGGCAACCCCGCGGACCAACATGCTGCGGCTC 138
Qy 142 TTAGTAAGAAGCTGGTATGCCACCTTTGTGATTGACTTTTTCARAGGAACCTAGCA 201
Db 139 GCCGGAAGAACTCGCCATCTGACCTGCTCGCGGACGTCGGCAAGGCGCTGTGCGG 198
Qy 202 AGCGTCTCGGATTAATTTTCATCTACAAGCGGTTTCTCTCTCATCTTTGGACTTTTG 261
Db 199 GTGCTGTGCGCGCTGGCTCGGCTCGCGTGATGGAGAGGCTGGGTCGGCATCGCC 258
Qy 262 GCTGTATCGGCCATACCTTCCTATCTTTGCGAGATTAAAGGTGGTAAGCTGTCCGA 321
Db 259 GCGGTGATCGGCCACTGTACCGCTGTACTTCAACTTCCGCGGCGCAAGGGTGTGCGC 318
Qy 322 ACCAGTGTGGAGTGAATTTTCG 344
Db 319 ACCGCGCGGCGCATGCTCTCGG 341

RESULT 11

US-09-815-242-9659

; Sequence 9659, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes In

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9659

; LENGTH: 612

; TYPE: DNA

; ORGANISM: Salmonella typhi

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(612)

US-09-815-242-9659

Query Match 9.0%; Score 57.8; DB 10; Length 612;
Best Local Similarity 48.9%; Pred. No. 1.5e-07;
Matches 155; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 23 TCTAGCCATCTGCTGGGTTCGATTCCTGCTGCGGATGACAAAGTATTTCTTC 82
Db 29 TCTTGGCTACCTCTCGGCTCAATTTCCAGTCCCAATCTGTGTTCCCGCATCGCCGGC 89
Qy 83 AAATCAATCTACCGGACGATGGTTCTGGTAACACTGGAACGACCAACACTTCGCGCATTT 142
Db 89 TGCCCGATCGCGAGAAAGGGGCTCCGGCAACCCCGCGGCGGACGATATTACGAATCG 148

Qy 143 TAGTAAGAAGCTGGTATGCAACCTTTGTGATTGACTTTTTCAAAGGAACCTAGCAA 202
Db 149 GTGGCAAGGAGCGCTCTCGCGGTACTTATTTTGGACATCTTAAAGGCATGTTGCCG 208
Qy 203 CGCTGCTTCCGATTAATTTTTCATCTACAAGCGGTTTCTCTCTCATCTTTGGACTTTTG 262
Db 209 TCTGGGCGCGTATGCTGTTAGCGGTACCCCTTTCTTGCTGGGCTTATCGCTATCGCG 268
Qy 263 CTGTATCGGCCATACCTTCCCTATCTTTGCGAGATTTAAAGGTGGTAAGCTGTGCGCA 322
Db 269 CTTGCTGGGACATCTGCGCGGTCTTTTGGCTTTTAAAGGCGGGAAGCGGTAGCAA 328
Qy 323 CCAGTGTCTGGAGTGATT 339
Db 329 CCGGCTTTGGCGCCATT 345

RESULT 12

US-09-815-242-3464/c

; Sequence 3464, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes In

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3464

; LENGTH: 189

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-3464

Query Match 8.8%; Score 56.6; DB 10; Length 189;
Best Local Similarity 62.2%; Pred. No. 1.9e-07;
Matches 89; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 243 TCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTGCGAGGATTTAA 302
Db 143 TTTAATTTGTTGCTTATTCGCTATCTATCTGACACGTTTATCTGTTATTTAAATTTCA 84
Qy 303 AGTGTTAAGGTGTCGCAACAGTGTGAGTGATTTTCGGATTTGCCCTATCTTCG 362
Db 83 AGTGCAAAAGCTGTTCGAACCTAGTGCAGGTGCTGCTTGGGAGTCAATCCGATCTTT 24
Qy 363 TCTCTACCTTGGGATTTATCTTCT 385
Db 23 ACTAATACTTGGCAATTTATCTTCT 1

RESULT 13

US-09-815-242-6206
; Sequence 6206, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6206
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(618)
US-09-815-242-6206

Query Match		8.3%	Score 53.6;	DB 10;	Length 618;
Best Local Similarity		48.1%	Pred. NO. 2.6e-06;		
Matches 152;		Conservative 0;	Mismatches 164;	Indels 0;	Gaps 0;
Qy	23	TCCTAGCCTATCTGCTGGGTTCCGATTCATCTGGTCTCTGGATTGGACAAAGTATCTTTC	82		
Db	29	TCATCGGTACCTCTGGCGCTCCATTTCCAGTGCCATCTGGTTTGGCGCTTGTGTGGGC	88		
Qy	83	AAATCAATCTACGGAGCATGGTCTTGTAACACTGGAACGACCAACACCTTCCGCAATT	142		
Db	89	TGCCGATCCGGAACACCGGCTCCGCAATCCAGGCGCAACCAATGTTCAGTATCG	148		
Qy	143	TAGGTAAAGAACTGGTATGGCAACCTTTGTGATGACTTTTCAAAGAACCCCTAGCAA	202		
Db	149	GTGGCAAGGAGGACCGCTAGCAGTACGATTTTCAGCGTTCTGAAGGAATTTGGCCG	208		
Qy	203	CGCTGCTCCGATTTATTTTCATCTACAAAGGCTTCTCCCTCATCTATCTTGGACTTTGG	262		
Db	209	TCGTGGGCGGTATGAATAGGTGTACGCCCTTCTTGCTAGGCTTAATTCGCATCGCG	268		
Qy	263	CTGTATTACGGCATACCTTCCCTATCTTTGCGAGGATTTAAAGGTGGTAAGGCTGTCGAA	322		
Db	269	CTGTCTTGGACACATCTGGCCCGTTTCTTCGGATTTAAAGGAGGAAAGGCGTGTGCTA	328		
Qy	323	CCAGTCTGGAGTAT	338		
Db	329	CCGCTTTTGGGCCAT	344		

RESULT 14

US-09-815-242-7616
; Sequence 7616, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7616
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(618)
US-09-815-242-7616

Query Match		6.0%	Score 38.4;	DB 10;	Length 618;
Best Local Similarity		45.4%	Pred. No. 0.082;		
Matches 138;		Conservative 0;	Mismatches 166;	Indels 0;	Gaps 0;
Qy	20	TATCTAGCCTATCTGCTGGGTTTCGATTCATCTGTTCTCTGGATTGGACAAAGTATCT	79		
Db	26	TACTCTCGCTACCTCTGCGGCTCCATTTCCAGCGCATTTCTGTTATGTCGCTTCTG	85		
Qy	80	TTCAAAATCAATCTACGCGGAGCATGGTCTGTTACACTGGAACGACCAACACCTTCCGCA	139		
Db	86	GCCTGCGCGACCCACGCGAGCAGCGGCTCAGCAACCCCGGGCCACTAACGCTACTAGAA	145		
Qy	140	TTTTAGGTAGAAAGCTGGTATGGCAACCTTTGTGTTGACTTTTCAAAGGAACTAG	199		
Db	146	TTGCGGAAAGGCGCGCGCTACGCGTACTTATTTTGTATGCTGAAGGCGATGCTCC	205		
Qy	200	CAAAGCTGCTCCGATTTATTTTCATCTACAAAGCGTTTCTCTCTCATCTTTTGGACTTT	259		
Db	206	CGGTGTGGGCGCTGCGGCTTACCGCGTTCTGCGGTTCTGCGGTTAGTGGCGATTG	265		
Qy	260	TGGCTGTATCGGCGCATACCTTCCCTATCTTTTCAGAGATTTAAAGTGGTAAGGCTGTCG	319		
Db	266	CCGCTCGTGGGCGCATATCTGGCCAGTCTTCTTTCATTTCCGCGCGGTAAGGCGCTCG	325		
Qy	320	CAAC	323		
Db	326	CCAC	329		

RESULT 15
US-09-938-842A-4649/c
; Sequence 4649, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4649
; LENGTH: 1118
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4649

Query Match	5.7%	Score 36.6	DB 9	Length 1118
Best Local Similarity	58.0%	Pred. No. 0.37		
Matches 83	Conservative 0	Mismatches 59	Indels 1	Gaps 1
QY	458	TTCTGCTCTTTCCACATTTTTGGTTTATCTCTGAGTAACATGACTCTCTCTTCATCGCTA	517	
Db	1020	TTCTGCTCTGTCCTCTTTAGTTTTTCTCTAAATATATTACTTTTGGTTCATCATTT	961	
QY	518	TTATCTTAGCACTTGCTAGTTTGATTATCATTCGTCATAGAAGAC - AATATAGTTCGATC	576	
		-		
Db	960	GAGTTCTCTCACTTGTCATATATTTTAAATAAACATAACTAAGAATATTGAACATATG	901	
QY	577	AAAAATAAAACTGAAAAATTTGGT	599	
Db	900	GAAACCCCAACTAACGATTAAGT	878	

Search completed: December 26, 2002, 00:32:36
Job time : 194.61 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:17:36 ; Search time 468.714 Seconds
(without alignments)
54.868 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLTALLIILAYLIGSIPGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	960	100.0	193	20	AA1980	AA1980	B. subtilis B-ynes
2	960	100.0	193	21	AA1981	AA1981	B. subtilis B-ynes
3	566.5	59.0	198	23	AB48775	AB48775	Listeria monocytog
4	476.5	49.6	204	23	AB40394	AB40394	Staphylococcus epi
5	468	48.8	213	21	AA192246	AA192246	S. pneumoniae S-yn
6	468	48.8	213	22	AAU37736	AAU37736	Streptococcus pneu
7	468	48.8	213	22	AAU37943	AAU37943	Streptococcus pneu
8	465.5	48.5	197	22	AAU32881	AAU32881	S. epidermidis ope
9	465.5	48.5	202	22	AAU33809	AAU33809	Staphylococcus aur
10	465.5	48.5	202	22	AAU36544	AAU36544	Staphylococcus aur

11	454	47.3	212	23	ABP26818	ABP26818	Streptococcus poly
12	453	47.2	240	23	ABP26819	ABP26819	Streptococcus poly
13	447	46.6	207	23	ABP30349	ABP30349	Streptococcus poly
14	445	46.4	213	23	ABP30349	ABP30349	Lactococcus lactis
15	275.5	28.7	114	20	AA1922561	AA1922561	Bacterial general
16	261.5	27.2	205	22	AAU36120	AAU36120	Klebsiella typhi c
17	257.5	26.8	203	22	AAU38164	AAU38164	Salmonella typhi c
18	255	26.6	199	22	AAU35404	AAU35404	Haemophilus influe
19	252	26.2	189	22	AAU36206	AAU36206	Pseudomonas aerugi
20	250.5	26.1	205	22	AAU34710	AAU34710	E. coli cellular p
21	230	24.0	220	22	AAU36021	AAU36021	Helicobacter pylor
22	228	23.8	262	19	AAU36021	AAU36021	H. pylori GHPO 108
23	228	23.8	262	19	AAU36021	AAU36021	Helicobacter pylor
24	226	23.5	173	21	AAU35860	AAU35860	Neisseria meningit
25	226	23.5	173	21	AAU35860	AAU35860	Neisseria meningit
26	224	23.3	173	21	AAU35860	AAU35860	Neisseria meningit
27	94.5	9.8	316	22	AAU44251	AAU44251	Propionibacterium
28	94	9.8	981	20	AAU34476	AAU34476	Porphorymonas ging
29	94	9.8	981	20	AAU34476	AAU34476	Porphorymonas ging
30	93.5	9.7	299	22	ABP26867	ABP26867	Human T2R01 amino
31	93	9.7	389	23	ABP26867	ABP26867	Lactococcus lactis
32	93	9.7	389	23	ABP26867	ABP26867	Staphylococcus epi
33	92.5	9.6	451	22	AAU34954	AAU34954	Enterococcus faeca
34	92.5	9.6	1014	22	AAU34954	AAU34954	C glutamicum prote
35	92	9.6	336	23	ABP26867	ABP26867	Streptococcus poly
36	91	9.5	783	22	AAU34954	AAU34954	C. glutamicum SRT
37	91	9.5	801	22	AAU34954	AAU34954	C. glutamicum prote
38	89.5	9.3	451	23	ABP29492	ABP29492	Streptococcus poly
39	89.5	9.3	463	22	AAU34954	AAU34954	S. epidermidis ope
40	89	9.3	155	21	AAU31973	AAU31973	Arabidopsis thalia
41	89	9.3	188	21	AAU31973	AAU31973	Arabidopsis thalia
42	89	9.3	251	21	AAU31964	AAU31964	Arabidopsis thalia
43	88.5	9.2	393	23	ABP26867	ABP26867	Streptococcus poly
44	88	9.2	248	22	AAU34954	AAU34954	Lactococcus lactis
45	88	9.2	254	23	AAU75790	AAU75790	S. epidermidis ope
							Human protein phos

ALIGNMENTS

RESULT 1
AA1922580
ID AA1922580 standard; Protein: 193 AA.
XX AA1922580;
AC AA1922580;
XX
XX 17-NOV-1999 (first entry)
DT
DE B. subtilis B-ynes protein sequence.
XX
XX General essential protein; pathogenic bacteria; pathogen; inhibitor;
KW bacterial growth; B-ynes.
XX
XX Bacillus subtilis.
OS
XX
XX WO9933871-A2.
PN
XX
XX 08-JUL-1999.
PD
XX
XX 30-DEC-1998; 98WO-US27918.
PF
XX
XX 31-DEC-1997; 97US-0070116.
PR
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX
XX Youngman P, Fritz C, Murphy C, Guzman L;
PI
XX
XX WPI; 1999-430230/36.
DR
XX
XX N-PSDB; AA20371.
PT
XX
XX Streptococcus pneumoniae general essential protein genes and proteins,
PT
XX
XX useful for identification of antibacterial agents.

PS Disclosure; Fig 24; 124pp; English.
XX
CC This sequence is the *Bacillus subtilis* B-ynes protein. B-ynes is
CC related to the *Streptococcus pneumoniae* general essential
CC protein (GEP) gene of the invention. The genes encoding the GEP
CC polypeptides are useful molecular tools for identifying similar genes in
CC pathogenic microorganisms, such as pathogenic strains of *Bacillus*. In
CC addition, the operons containing genes encoding GEP and the polypeptides
CC themselves, are useful targets for identifying compounds that are
CC inhibitors of the pathogens in which the GEP are expressed. Such
CC inhibitors are useful for inhibiting bacterial growth by being
CC bacteriostatic or bacteriocidal.
XX

SO Sequence 193 AA;
Query Match 100.0%; Score 960; DB 20; Length 193;
Best Local Similarity 100.0%; Pred. No. 9e-100;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLIALLIILAYLIGSIPSGLVKLGAKGIDIREHSGNLCATNAFTLGVKAGSVVIAGD 60
DB 1 MLIALLIILAYLIGSIPSGLVKLGAKGIDIREHSGNLCATNAFTLGVKAGSVVIAGD 60
QY 61 ILKGTALATLPFLMHVDIHPHLLAGVFAVLGHVFPPIFAKFGKRAVATSGVLLFYAPLLF 120
DB 61 ILKGTALATLPFLMHVDIHPHLLAGVFAVLGHVFPPIFAKFGKRAVATSGVLLFYAPLLF 120
QY 121 ITWVAVFFIPLYLTKFVLSLSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
DB 121 ITWVAVFFIPLYLTKFVLSLSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
QY 181 RIINKTEPKVKWL 193
DB 181 RIINKTEPKVKWL 193

RESULT 2
AAY92247
ID AAY92247 standard; Protein; 193 AA.
XX
AC AAY92247;
XX
DT 10-AUG-2000 (first entry)
XX
DE B. subtilis B-ynes polypeptide.
XX
KW B-ynes; survival; antibacterial; inhibitor.
XX
OS *Bacillus subtilis*.
XX
PN WO200020627-A1.
XX
PD 13-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22665.
XX
PR 30-SEP-1998; 98US-0163445.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Fritz C, Youngman P, Guzman L;
XX
DR WPI; 2000-303799/26.
XX
DR N-PSDB: AAA09182.
XX
PT Methods for identifying an antibacterial agent for treating
PT *Streptococcus pneumoniae* infections comprises detecting an interaction
PT between a yneS polypeptide and a test compound
XX
PS Disclosure; Fig 2; 65pp; English.
XX
CC *Bacillus subtilis* yneS gene is a homologue of *Streptococcus pneumoniae*
CC yneS gene, which encodes a polypeptide (S-ynes) essential for survival

CC for a wide range of bacteria. Identifying an antibacterial agent
CC comprises contacting a yneS polypeptide (S-ynes) with a test compound and
CC detecting an interaction of the test compound with the S-ynes polypeptide
CC which indicates that the compound is an antibacterial agent.
CC Alternatively, detecting a decrease in function of the polypeptide
CC contacted with the test compound and determining whether the compound
CC inhibits growth of bacteria, relative to the growth of bacteria cultured
CC in the absence of a test compound where inhibition of growth indicates
CC the compound is an antibacterial agent. Inhibitors of S-ynes function
CC are useful for treating a *Streptococcus pneumoniae* infection in mammals.
XX

SO Sequence 193 AA;
Query Match 100.0%; Score 960; DB 21; Length 193;
Best Local Similarity 100.0%; Pred. No. 9e-100;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLJALLIILAYLIGSIPSGLVKLGAKGIDIREHSGNLCATNAFTLGVKAGSVVIAGD 60
DB 1 MLJALLIILAYLIGSIPSGLVKLGAKGIDIREHSGNLCATNAFTLGVKAGSVVIAGD 60
QY 61 ILKGTALATLPFLMHVDIHPHLLAGVFAVLGHVFPPIFAKFGKRAVATSGVLLFYAPLLF 120
DB 61 ILKGTALATLPFLMHVDIHPHLLAGVFAVLGHVFPPIFAKFGKRAVATSGVLLFYAPLLF 120
QY 121 ITWVAVFFIPLYLTKFVLSLSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
DB 121 ITWVAVFFIPLYLTKFVLSLSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
QY 181 RIINKTEPKVKWL 193
DB 181 RIINKTEPKVKWL 193

RESULT 3
ABB48775
ID ABB48775 standard; Protein; 198 AA.
XX
AC ABB48775;
XX
DT 05-FEB-2002 (first entry)
XX
DE *Listeria monocytogenes* protein #1479.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS *Listeria monocytogenes*.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) ENST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX
XX Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and
PT related polypeptides

PT Streptococcus pneumoniae infections comprises detecting an interaction
 PT between a ynes polypeptide and a test compound
 PS Disclosure: Fig 1; 65pp; English.
 XX
 XX Streptococcus pneumoniae ynes gene encodes a polypeptide (S-ynes)
 CC essential for survival for a wide range of bacteria. Identifying an
 CC antibacterial agent comprises contacting a ynes polypeptide (S-ynes) with
 CC a test compound and detecting an interaction of the test compound with
 CC the S-ynes polypeptide which indicates that the compound is an
 CC antibacterial agent. Alternatively, detecting a decrease in function of
 CC the polypeptide contacted with the test compound and determining whether
 CC the compound inhibits growth of bacteria, relative to the growth of
 CC bacteria cultured in the absence of a test compound where inhibition of
 CC growth indicates the compound is an antibacterial agent. Inhibitors of
 CC S-ynes function are useful for treating a Streptococcus pneumoniae
 CC infection in mammals.
 XX
 XX Sequence 213 AA;
 SQ
 Query Match 48.8%; Score 468; DB 21; Length 213;
 Best Local Similarity 48.0%; Pred. No. 2.1e-44;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLIALLLIAYLGISPSGLIVGKLAGIDIREHSGNLGNATNFRILGKAGSVVIAGD 60
 DB 1 MITIVLLIAYLGISPSGLIWIGVFFQINLREHSGNGTGTNTFRILGKAGMATFVID 60
 QY 61 ILKGTALATPLMHVD-IHPLLAGVFAVLGHVPIFAKFGKAVATSGVLLFYAPLL 119
 DB 61 FFKGTALATPLPIIFHLQGVSPLEGLLAVIGHTFTPIFAGFGKAVATSGVIFGPAFIF 120
 QY 120 FITWAVFFIYLYTKFVLSMLTGIVTI-----YSFFV--HDTYLLIVVTLTIF 170
 DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLPFLFGFILSNYDSLFTAILALASL 180
 QY 171 VIYHRANKRIINKTEPKVKW 192
 DB 181 IIRHKDNRIARKNTENLVPM 202
 RESULT 6
 AAU37736
 ID AAU37736 standard; Protein; 213 AA.
 AC
 AC AAU37736;
 DT 14-FEB-2002 (first entry)
 DE Streptococcus pneumoniae cellular proliferation protein #165.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX Streptococcus pneumoniae.
 OS
 XX WO200170955-A2.
 PN 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 PF 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2001; 2000US-191078P.
 XX 23-MAY-2000; 2000US-206948P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS55595.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX Example 3; Seq ID No 13329; 51pp; English.
 PS
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 213 AA;
 SQ
 Query Match 48.8%; Score 468; DB 22; Length 213;
 Best Local Similarity 48.0%; Pred. No. 2.1e-44;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLIALLLIAYLGISPSGLIVGKLAGIDIREHSGNLGNATNFRILGKAGSVVIAGD 60
 DB 1 MITIVLLIAYLGISPSGLIWIGVFFQINLREHSGNGTGTNTFRILGKAGMATFVID 60
 QY 61 ILKGTALATPLMHVD-IHPLLAGVFAVLGHVPIFAKFGKAVATSGVLLFYAPLL 119
 DB 61 FFKGTALATPLPIIFHLQGVSPLEGLLAVIGHTFTPIFAGFGKAVATSGVIFGPAFIF 120
 QY 120 FITWAVFFIYLYTKFVLSMLTGIVTI-----YSFFV--HDTYLLIVVTLTIF 170
 DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLPFLFGFILSNYDSLFTAILALASL 180
 QY 171 VIYHRANKRIINKTEPKVKW 192
 DB 181 IIRHKDNRIARKNTENLVPM 202
 RESULT 7
 AAU37943
 ID AAU37943 standard; Protein; 213 AA.
 AC
 AC AAU37943;
 DT 14-FEB-2002 (first entry)
 DE Streptococcus pneumoniae cellular proliferation protein #372.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX Streptococcus pneumoniae.
 OS
 XX WO200170955-A2.
 PN 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS55802.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 XX Example 3; Seq ID No 13536; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 213 AA;
 Query Match 48.8%; Score 468; DB 22; Length 213;
 Best Local Similarity 48.0%; Pred. No. 2.1e-44;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLTALLILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTGLGKAGSVVIAGD 60
 DB 1 MTIVLLILAYLLGSPGLWIGQVFFOINLREHSGNGTNTNTRILCKKAGMATFVID 60
 QY 61 ILKGTALATLPLMHVD-IHPLLACGVFAVLGHVFPFIFAKFGKGAATSGGVLLFYAPLL 119
 DB 61 PFKGTLATLPIIFHLQGVSPILFGLLAVIGTTPFIFAGFGKGAATSGAGVIFGAPF 120
 QY 120 FITWAVVFIFLYLTKFVLSLSMLTGIVTI-----YSFV--HDYLLIVVTLTIF 170
 DB 121 CLYLAIIFEGALYLSMSLSSTASIAAVIGLVLLFPLFGFTLSNYDSLFTAILALASL 180
 QY 171 VIYHRANKRIINKTEPKVKV 192
 DB 181 IIRHKDNRIARKNTENLVPW 202
 XX
 XX AAG82881
 ID AAG82881 standard; Protein; 197 AA.
 AC AAG82881;
 XX
 XX 03-SEP-2001 (first entry)
 XX

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2856.
 XX
 KW *Staphylococcus epidermidis* SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX
 OS *Staphylococcus epidermidis*.
 XX
 PN W0200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 XX 09-NOV-2000; 2000WO-US30782.
 PF
 XX
 XX 09-NOV-1999; 99US-0164258.
 XX
 XX (GLAX) GLAXO GROUP LTD.
 PA
 XX Kimmerly WJ;
 PI
 XX WPI; 2001-316495/33.
 DR N-PSDB; AAH53731.
 XX
 XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 PT useful for vaccinating against infections, e.g. endocarditis -
 PT
 XX Claim 13; Page 745; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC *S. epidermidis* polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 XX Sequence 197 AA;
 Query Match 48.5%; Score 465.5; DB 22; Length 197;
 Best Local Similarity 48.7%; Pred. No. 3.7e-44;
 Matches 96; Conservative 35; Mismatches 57; Indels 9; Gaps 2;
 QY 6 LIIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTGLGKAGSVVIAGDILKGT 65
 DB 1 MLILSYLCAPFSGLLIICKLFKKDIROYGSGNTGATNSFRVLGRPAGFVTFPIKGF 60
 QY 66 LATALP--FLMHVD-----IHPLLACGVFAVLGHVFPFIFAKFGKGAATSGGVLLFYA 116
 DB 61 ITVFPFLWFPVHADVISTFTFTNGLVGLFALGHVYPIYLFKNGKGAATSGAVVLGVN 120
 QY 117 PLLFITWAVVFIFLYLTKFVLSLSMLTGIVTVYVSFFVHDTYLLIVVTLTIFVIYRHR 176
 DB 121 PILLLILAIIFFSVLKIFKVFYLSLSIIAISCVIGSIIIHDIYLLAVSGIVSIIILIRHK 180
 QY 177 ANIKRIINKTEPKVKWL 193
 DB 181 SNIVRIFKGEPEPKIKMW 197
 XX
 XX RESULT 9
 XX AAU33809

AAU33809 standard; Protein; 202 AA.
AAU33809;
14-FEB-2002 (first entry)
Staphylococcus aureus cellular proliferation protein #85.
Antisense; prokaryotic cellular proliferation protein;
antibiotic; antibacterial; drug design.
Staphylococcus aureus.
WO200170955-A2.
27-SEP-2001.
21-MAR-2001; 2001WO-US09180.
21-MAR-2000; 2000US-191078P.
23-MAY-2000; 2000US-206848P.
26-MAY-2000; 2000US-207727P.
23-OCT-2000; 2000US-242578P.
27-NOV-2000; 2000US-253625P.
22-DEC-2000; 2000US-257931P.
16-FEB-2001; 2001US-269308P.
(ELIT-) ELITRA PHARM INC.
Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
Yamamoto RT, Xu HH;
WPI: 2001-611495/70.
N-PSDB; AAS51668.
New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Example 3; Seq ID No 5305; 51lpp; English.
The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the
genes, their use in the discovery of novel antibiotics, the essential
genes themselves and the encoded proteins. The prokaryotes used are
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
invention is also useful for the identification of potential new targets
to identify proteins used in proliferation, to express these proteins,
and to obtain antibodies capable of binding to the expressed proteins.
The proteins can be used to screen compounds in rational drug discovery
programmes. The antisense nucleic acid sequence is also useful to screen
for homologous nucleic acids which are required for cell proliferation in
essential prokaryotic cellular proliferation protein.
Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic
format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.
Query Match 48.5%; Score 465.5; DB 22; Length 202;
Best Local Similarity 46.5%; Pred. No. 3.8e-44;
Matches 94; Conservative 40; Mismatches 59; Indels 9; Gaps 2;
OY 1 MLIALLIILAYIGTSPGLVIGKLAGIDIREHSGNLTGATNAPRTLCVKGAGSVVIAGD 60
DB 1 MMLIYLLSYLIGFPSPGVIGKLFKKDKIQFGSGTGNATNSFRVLCRPGAFVLTFLD 60
OY 61 ILKGTALATAPFLM--HVD-----IHPLLAGVFAVLGHVFPPIPAKFGKAVATSGV 111
DB 61 IFKGFITVFPFLPVLPHADGPISFTFTNGLIVGLFAILGHVYVPYLVKFGGKAVATSGV 120

OY 112 LLEYAPLLFITMVAVFFIFLYLTKEVSLSSMLTGITYTVIYFFVHDYTLIIWVLLTIFV 171
DB 121 VLGWNPILLILLALIIFFIVLKIFKYSLSASIVAAICCVIGSLTIQDYILLVWSFLVSL 180
OY 172 IYRHRANIKRIINTEPKVKWL 193
DB 181 IYRHRNSISRIFRGEPEAKWM 202
RESULT 10
AAU36544
ID AAU36544 standard; Protein; 202 AA.
XX AAU36544;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #714.
XX
KW Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
DR N-PSDB; AAS51668.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5305; 51lpp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 202 AA;

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS.
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 240 AA;
 Query Match 47.28; Score 453; DB 23; Length 240;
 Best Local Similarity 47.08; Pred. No. 1.2e-42;
 Matches 95; Conservative 34; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLIALLIILAYLIGSIPGLIVGKLAKGIDIREHSGNGLGATNAPRTTGVKAGSVVIAGD 60
 DB 28 MKLLFTIYIAYLIGSIPGLIGQVYFHNIREHSGNGTNTTFRILGVKAGTATLAID 87
 QY 61 ILKGTALPFLMHV-DIHPLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLL 119
 DB 88 MFKGTSLILPIIFOMTSISSIAIGFFAVLGHVFFIFANFKGKAVATSGVLLGFAPLY 147
 QY 120 FITWAVAFIFLYLTKFVLSLM---LTGIYTVIYSFFVH-----DYLLIVVTLTIF 170
 DB 148 LFFLASIEVLVLYFSMISLASVSAIVGVLVLTFFPAIHLLPNYDYFLTIFVILLAFI 207
 QY 171 VYHRANIKRIINKTEPKVW 192
 DB 208 IIRHKONISRIKHHTENLIPW 229
 RESULT 13
 ABP30349
 ID ABP30349 standard; Protein; 207 AA.
 XX
 AC ABP30349;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 9874.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Massignani V, Margarit Ros Yi, Grandi G, Fraser C;
 PI Tettelin H;
 PI
 XX WPI; 2002-352536/38..

DR N-PSDB; ABN70980.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1; Page 4116; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 207 AA;
 Query Match 46.68; Score 447; DB 23; Length 207;
 Best Local Similarity 48.48; Pred. No. 4.7e-42;
 Matches 93; Conservative 35; Mismatches 54; Indels 10; Gaps 3;
 QY 6 LILAYLIGSIPGLIVGKLAKGIDIREHSGNGLGATNAPRTTGVKAGSVVIAGDILKGT 65
 DB 1 MILLIAYLIGSIPGLIGQVYFHNIREHSGNGTNTTFRILGVKAGITVLTIDILKGT 60
 QY 66 LATALPFLMHV-DIHPLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLLFTMV 124
 DB 61 LATLIPIILGITTVSPFFIGFAIIGHTEPIFAQFKGKAVATSGVLLGFAPSFYLL 120
 QY 125 AVFFIELYLTFRVLSLM---LTGIYTVIYSFFV-----HDTYLLIVVTLTIFVIYRH 175
 DB 121 VIFLLTYLFSMISLSSITVAVVGLSVLIPPLVGLFILTIDYDIFTTIVVILMALTIIIRH 180
 QY 176 RANIKRIINKTE 187
 DB 181 QDNIKRIKROE 192
 RESULT 14
 ABB54294
 ID ABB54294 standard; Protein; 213 AA.
 XX
 AC ABB54294;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein ykac.
 XX
 KW Lactococcus lactis protein ykac.
 XX
 OS Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 PN Lactococcus lactis IL1403.
 XX
 PD FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: December 26, 2002, 00:31:11 ; Search time 300.909 Seconds
(without alignments)
61.660 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALIILAYLIGSIPSGL.....RHRANKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	960	100.0	193	1	A69892	conserved hypothet
2	568.5	59.2	198	2	A81598	conserved hypothet
3	566.5	59.0	198	2	AD1235	B. subtilis ynes p
4	468	48.8	213	2	C95098	conserved hypothet
5	468	48.8	213	2	C97966	conserved hypothet
6	464.5	48.4	202	2	A99911	conserved hypothet
7	445	46.4	213	2	B86747	conserved hypothet
8	352.5	36.7	222	1	S75137	hypothetical prote
9	339.5	35.4	226	2	AC1868	hypothetical prote
10	333	34.7	202	2	A82737	conserved hypothet
11	333	34.7	205	2	H97517	hypothetical prote
12	322	33.5	206	2	G83854	hypothetical prote
13	318.5	33.2	239	2	S73812	hypothetical prote
14	309.5	32.2	196	2	C72253	conserved hypothet
15	302	31.5	192	2	E70359	conserved hypothet
16	301	31.4	201	2	AC3593	probable integral
17	299	31.1	208	2	B82313	conserved hypothet
18	296	30.8	198	2	A75324	conserved hypothet
19	286	29.8	216	2	A80080	probable membrane
20	284.5	29.6	218	2	E87552	conserved hypothet
21	284.5	29.6	239	1	C64227	hypothetical prote
22	271.5	28.3	198	2	C97208	uncharacterized co
23	270.5	28.2	257	2	A82894	conserved hypothet
24	263.5	27.4	202	2	D81378	probable integral
25	261.5	27.2	224	2	G90548	conserved hypothet
26	257.5	26.8	203	2	AF0892	probable membrane
27	257	26.8	200	2	C81894	probable integral
28	257	26.8	200	2	H81126	conserved hypothet
29	255	26.6	199	2	E64146	hypothetical prote

30	252	26.2	189	2	A83573	conserved hypothet
31	250.5	26.1	205	1	A65094	ylGH protein - Esc
32	250.5	26.1	205	2	F91121	hypothetical prote
33	250.5	26.1	205	2	E85966	hypothetical prote
34	230	24.0	220	2	B71811	hypothetical prote
35	228	23.8	262	1	E64708	conserved hypothet
36	196.5	20.7	203	2	F72273	conserved hypothet
37	165	17.2	195	2	A75295	conserved hypothet
38	141	14.7	559	2	B75477	conserved hypothet
39	115	12.0	463	2	A69905	conserved hypothet
40	110.5	11.5	421	2	G72300	conserved hypothet
41	104.5	10.9	963	2	AF2119	hypothetical prote
42	100	10.4	374	2	F90361	transport membrane
43	98.5	10.3	276	2	A75037	hypothetical prote
44	97.5	10.2	322	2	F95970	probable sugar upt
45	97.5	10.2	351	1	C69198	phospho-N-acetylmu

ALIGNMENTS

RESULT 1
A69892
conserved hypothetical protein ynes - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69892
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hullo
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Sc
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrat, P.; Togononi, A.; Tosato, V.; Uch
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus sub
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69892
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-193 <KUN>
A:Cross-references: GB:Z99113; GB:AL009126; MID:g2634090; PIDN:CAB13690.1; PID:g26
A:Experimental source: strain 168
C:Genetics:
A:Gene: ynes
C:Superfamily: Escherichia coli yglH protein

Query Match	100.0%	Score 960;	DB 1;	Length 193;
Best Local Similarity	100.0%	Pred. No. 5.2e-69;		
Matches 193;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MLIALIILAYLIGSIPSGLVGKLA	GIDIREHSGNLTGATNAPRTLC	GVKAGSVVIAGD 60
Db	1	MLIALIILAYLIGSIPSGLVGKLA	GIDIREHSGNLTGATNAPRTLC	GVKAGSVVIAGD 60
Qy	61	ILKGTALATPLMHVDHPLLAGV	FAVLGHVFPIFAFKGKAVATSG	VLLFYAPLLF 120
Db	61	ILKGTALATPLMHVDHPLLAGV	FAVLGHVFPIFAFKGKAVATSG	VLLFYAPLLF 120
Qy	121	ITWAVFFIFLYLTFTVSLSSML	TGIYTVIYVSFFVHDVTLIVV	TLLTIFVYRHRANIK 180
Db	121	ITWAVFFIFLYLTFTVSLSSML	TGIYTVIYVSFFVHDVTLIVV	TLLTIFVYRHRANIK 180
Qy	181	RIINKTEPKVKWL	193	
Db	181	RIINKTEPKVKWL	193	

```
RESULT 2
AB1598
conserved hypothetical protein, B. subtilis ynes protein homolog lin1323 [imported] - L
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96554.1; PID:gl6413796; GSPDB:GN00178
A:Experimental source: strain Clilp1262
C:Genetics:
A:Gene: lin1323
C:Superfamily: Escherichia coli ygiH protein

Query Match 59.2%; Score 568.5; DB 2; Length 198;
Best Local Similarity 57.7%; Pred. No. 4.7e-38;
Matches 112; Conservative 35; Mismatches 44; Indels 3; Gaps 1;

Qy 3 IALLIILAYLIGSIPSLVGLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 62
Db 5 LILLSLLAYVIGSIPSLVGLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 64

Qy 63 KGTALATLPEL--HHVDIH-PLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLL 119
Db 65 KGTATLPLFFFLQNVDFHFWLLTGAFALIGHSPFLFAGFGKAVATSGGVLLFYAPLL 124

Qy 120 FITWAVFFILYLTKFVSLSSMLTGIYTVIYFFVHDYTLIVVTLTIFVIYHRANI 179
Db 125 FVAALVVFLLTKISKYVSLSSMLGALAALISFFMGDWILIVLACIALFVIRHRANI 184

Qy 180 KRIINKTEPKVKWL 193
Db 185 TRIRNGEPEKIKWM 198

RESULT 3
AD1235
B. subtilis ynes protein homolog lml284 [imported] - Listeria monocytogenes (strain EGD
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1235
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99362.1; PID:gl6410700; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lml284
C:Superfamily: Escherichia coli ygiH protein

Query Match 59.0%; Score 566.5; DB 2; Length 198;
Best Local Similarity 58.8%; Pred. No. 6.8e-38;
Matches 114; Conservative 33; Mismatches 44; Indels 3; Gaps 2;
```

```
Qy 3 IALLIILAYLIGSIPSLVGLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 62
Db 5 LILLSLLAYVIGSIPSLVGLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 64

Qy 63 KGTALATLPEL--HHVDIH-PLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLL 119
Db 65 KGTATLPLFFFLQNVDFHFWLLTGAFALIGHSPFLFAGFGKAVATSGGVLLFYAPLL 124

Qy 120 FITWAVFFILYLTKFVSLSSMLTGIYTVIYFFVHDYTLIVVTLTIFVIYHRANI 179
Db 125 FVAALVVFLLTKISKYVSLSSMLGALAALISFFMGDWILIVLACIALFVIRHRANI 184

Qy 180 KRIINKTEPKVKWL 193
Db 185 TRIRNGEPEKIKWM 198

RESULT 4
C95098
conserved hypothetical protein SP0851 [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95098
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.;
nson, T.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaj
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Mori
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74980.1; PID:gl4972323; GSPDB:GN00164; TI
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0851
C:Superfamily: Escherichia coli ygiH protein

Query Match 48.8%; Score 468; DB 2; Length 213;
Best Local Similarity 48.0%; Pred. No. 4.4e-30;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

Qy 1 MLIALIILAYLIGSIPSLVGLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGD 60
Db 1 MITIVLLIILAYLIGSIPSLVGLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGD 60

Qy 61 ILKGTALATLPELHVD--IHPLLAVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLL 119
Db 61 FFKGTATLPLFFFLQNVDFHFWLLTGAFALIGHSPFLFAGFGKAVATSGGVLLFYAPLL 120

Qy 120 FITWAVFFILYLTKFVSLSSMLTGIYTVI-----YSPFV--HDTYLLIVVTLTIF 170
Db 121 CLYLAIIFGALYGLSMISLSVTSIAIAVIGVLLFPLFGFILSNYDSLTAILALASL 180

Qy 171 VIYHRANIKRIINKTEPKVKW 192
Db 181 IIRHKDNIARIKKNKTNLVPW 202

RESULT 5
C97966
conserved hypothetical protein spr0755 [imported] - Streptococcus pneumoniae (stra
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97966
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren,
J. P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas,
```


Db 66 ISKGVMAVALVRAIYSGDWLPAALPAQWNLTLGVIAVVLGHSHKSPFLKFSGGKSVATS 125
Qy 109 GGVLLFYAPLLFITWVAFFFLYLVKFSVSSMLTGIYT--VIYSFFVHDYLLIVVTL 166
Db 126 LGVPLMLNLWALGTFLATFLVIFTRIVSSSIVAALAVNGIALALQPPYLAFTF-L 184
Qy 167 LTFIVYIYHRANIKRIINKTEPKV 190
Db 185 AGMYVIVRHRNTNRILOQTEPKL 208

RESULT 9

AC1868
hypothetical protein all0492 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC1868
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriouchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1868
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-226 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072450.1; PID:gl7129837; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Gene: all0492
C:Superfamily: Escherichia coli ygiH protein

Query Match 35.4%; Score 339.5; DB 2: Length 236;
Best Local Similarity 40.2%; Pred. No. 6.7e-20;
Matches 84; Conservative 37; Mismatches 87; Indels 21; Gaps 5;
Qy 2 LIALLIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTLVGVAGSVVIAGDI 61
Db 7 LCGAVLVAYLLGSPPTGYIAVKQLKIDIREVSGSGTATNVLRTLGKGPFAVLGLDC 66
Qy 62 LKGTATLAL-----PFLMHVDI-HP---LLAGVFAVLGHVFFIFAKFKGCK 103
Db 67 LKGVLAIALVYLFASFASQNLIPPTVNIELQWPLVTLGAILGHKSFIFLGTGCK 126
Qy 104 AVATSGGVLLFYAPLLFITWVAFFFLYLVKFSVSSMLTGIYTIVYFFVHD--TYLL 161
Db 127 SVATSLGILLANWQVGLATGCVFAVVAISRVLSLSSIMGAIVSVVFLQQLPYIL 186
Qy 162 IVVTLTIFVIYHRANIKRIINKTEPKV 190
Db 187 FGIA-GGLYVILHRHSNRIERLAGTEPKI 214

RESULT 10

AB2737
conserved hypothetical protein Atul306 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB2737
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL42312.1; PID:gl7739714; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul306
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ygiH protein

Query Match 34.7%; Score 333; DB 2: Length 202;
Best Local Similarity 38.1%; Pred. No. 2e-19;
Matches 72; Conservative 43; Mismatches 74; Indels 0; Gaps 0;

Qy 2 LIALLIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTLVGVAGSVVIAGDI 61
Db 10 LLAALIGLIGLIGSPFGLILTRMAGLGDVVRKIGSGNIGATNVLRTGNKLAATLLDA 69
Qy 62 LKGTATLALPFLMHVDIHPHLLAGVFAVLGHVFFIFAKFKGKAVATSGVLLFYAPLLFI 121
Db 70 LKGTAAVLVANALNGYEASLVAGFFAFGLHLPVWLFGKGVAVVIGVLLGAAPLML 129
Qy 122 TMVAVFFIFLYLVKFSVSSMLTGIYTIVYFFVHDYLLIVVTLTIFVIYHRANIKR 181
Db 130 AFALIWLTATFITRYSSLSALLAMLIIPVALWVLGPEKTAMLVTLSSVISWKKHREIRR 189
Qy 182 IINKTEPKV 190
Db 190 LMACTESRI 198

RESULT 11

H97517
hypothetical protein AGR_C_2402 [imported] - Agrobacterium tumefaciens (strain C58).
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97517
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Go:
A: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markel:
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacteri
A:Reference number: A97359; PMID:11743194
A:Accession: H97517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87097.1; PID:gl5156359; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2402
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ygiH protein

Query Match 34.7%; Score 333; DB 2: Length 205;
Best Local Similarity 38.1%; Pred. No. 2e-19;
Matches 72; Conservative 43; Mismatches 74; Indels 0; Gaps 0;

Qy 2 LIALLIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTLVGVAGSVVIAGDI 61
Db 13 LLAALIGLIGLIGSPFGLILTRMAGLGDVVRKIGSGNIGATNVLRTGNKLAATLLDA 72
Qy 62 LKGTATLALPFLMHVDIHPHLLAGVFAVLGHVFFIFAKFKGKAVATSGVLLFYAPLLFI 121
Db 73 LKGTAAVLVANALNGYEASLVAGFFAFGLHLPVWLFGKGVAVVIGVLLGAAPLML 132
Qy 122 TMVAVFFIFLYLVKFSVSSMLTGIYTIVYFFVHDYLLIVVTLTIFVIYHRANIKR 181
Db 133 AFALIWLTATFITRYSSLSALLAMLIIPVALWVLGPEKTAMLVTLSSVISWKKHREIRR 192
Qy 182 IINKTEPKV 190
Db 193 LMACTESRI 201

RESULT 12

G83854
hypothetical protein BH1639 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83854
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83854
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AF001512; GB:BA000004; NID:g10174030; PIDN:BA05358.1; GSPDB:GN00
C:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1639
C:Superfamily: Escherichia coli ygiH protein

Query Match 33.5%; Score 322; DB 2; Length 206;
Best Local Similarity 37.3%; Pred. No. 1.5e-18;
Matches 75; Conservative 43; Mismatches 61; Indels 22; Gaps 4;

Qy 5 LLIILAYLIGSPGLIVGKLAGIDIREHSGNLTGATNAPRTLVGKAGSVVIAGDILKG 64
Db 6 LLIIVGSLGVSFSYIIIAIKKIDIRHSGNAGATNLTVLGVGPAVTVLLDILKG 65

Qy 65 TLATLALPFLMHVDIHP-----LLAGVFAVLGHVFPIFAKFGKAVATSGVLLFYAP 117
Db 66 VIATV-----VTVLTDPDGGWFAAAGIAIIGHNWIYFGKGGKAVTTIGVLASLVP 121

Qy 118 LLEFI-----TWAVFFIFLYLKFSLSMLTGIVTVISFFVHDTYLLIVVTLTI 169
Db 122 LAALVAGTAVIGSIWTVRSVLSGLLFTVLTALLAVLSQWFGYPVAVIYLLTIIVAIL- 180

Qy 170 FVIYHRANIKRIINKTEPKV 190
Db 181 --MWRHRSIQRLSCTENKL 199

RESULT 13
S73812
hypothetical protein ygiH - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein H91_orf239
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73812
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-239 <HIM>
A:Cross-references: EMBL:AE000047; GB:U00089; NID:g1674162; PIDN:AAB96134.1; PID:g167417
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: ygiH
A:Genetic code: SGC3
C:Superfamily: Escherichia coli ygiH protein

Query Match 33.2%; Score 318.5; DB 2; Length 239;
Best Local Similarity 34.2%; Pred. No. 3.2e-18;
Matches 79; Conservative 45; Mismatches 68; Indels 39; Gaps 6;

Qy 1 MLTALLIILAYLIGSPGLIVGKLAGIDIREHSGNLTGATNAPRTLVGKAGSVVIAGD 60
Db 9 LLIIVSLVIGLWGSVMFADVFGKIILNK-DVFKLSKNPGATNSIRVGLKIGFLVGLCD 67

Qy 61 ILKGTATLALPFLM-----HVDIH-----PLLAGVFAVLGHVFPIFAKFGKAV 105
Db 68 ALKGLFAVFSFLTSFWMLOQLVNNQYKQVYLYTSLSCFAATIGHIFLVPKFGKRAI 127

Qy 106 ATSGVLLFYAPLFFITMVAVFIFLYLTKEVSLSSMLTGIIYTVI-----YSFFV-- 155
Db 128 ATTGCSLLAISLWFWFICLVLLWLLVTLTKYVSLASLVFFILAILIILVPHLDLYFPKP 187

Qy 156 -----HDTYLLIVVTL-----LTFIVYHRANIKRIINKTEPKVKWL 193
Db 188 NPINAISYQNDWYIILFEVLWYWPULTIAVFWLHRKNIHRLNKTENKVTQL 238

RESULT 14
C72253
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: C72253
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.;
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardso
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from geno
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: C72253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <ARN>
A:Cross-references: GB:AE001796; GB:AE000512; NID:g4982004; PIDN:AAD36515.1; PID:g
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1447
C:Superfamily: Escherichia coli ygiH protein

Query Match 32.2%; Score 309.5; DB 2; Length 196;
Best Local Similarity 37.6%; Pred. No. 1.4e-17;
Matches 71; Conservative 34; Mismatches 79; Indels 5; Gaps 3;

Qy 5 LLIILAYLIGSPGLIVGKLAGIDIREHSGNLTGATNAPRTLVGKAGSVVIAGDILKG 64
Db 5 LFPILGYFIGSIPSPYLIPLKWLKIDVRKVGSGNVTATRTTGPVGGICLLLDALKG 64

Qy 65 TLATLALPFLMHVDIHP-LAGVFAVLGHVFPIFAKFGKAVATSGVLLFYAPLFFITM 123
Db 65 FFPVFTITFGSDKISVLSLTATVGLHDFIPFMKFGKGVASTGLIIFCLSNPTGLVF 124

Qy 124 VAVFFIFLYLKFSLSMLTGIVTVISFFV--HDTYLLIVVTLTTFIVYHRANIKR 181
Db 125 TLTMLVIVMLTKYASLSGLVALYVSALLGYLLKGYDTGMLELI--LAVLSTLRHSENIQR 182

Qy 182 IINKTEPKV 190
Db 183 LLNGTERKV 191

RESULT 15
E70359
conserved hypothetical protein aq_676 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C:Accession: E70359
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-192 <AQF>
A:Cross-references: GB:AE000701; NID:g2983260; PIDN:AAC08869.1; PID:g2983275; GB:A
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_676
C:Superfamily: Escherichia coli ygiH protein

Search completed: December 26, 2002, 01:27:06
Job time : 301.909 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:27:20 ; Search time 137.382 Seconds
(without alignments)
58.268 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALIILAYLIGSIPSGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	960	100.0	193	1	YNES_BACSU		Q45064 bacillus su
2	468	48.8	213	1	Y851_STRPN		Q54916 streptococc
3	352.5	36.7	222	1	Y373_SYNY3		P73933 synechocyst
4	318.5	33.2	239	1	Y247_MYCPN		P73428 mycoplasma
5	302	31.5	192	1	Y676_AQUAE		O66905 aquifex aeo
6	284.5	29.6	239	1	Y247_MYCGE		P47489 mycoplasma
7	255	26.6	199	1	Y4603_HAEMOPHILUS		P44603 haemophilus
8	250.5	26.1	205	1	YGIH_ECOLI		P31056 escherichia
9	230	24.0	220	1	YF09_HELPJ		Q92jbl helicobacte
10	228	23.8	220	1	YF09_HELPJ		Q26039 helicobacte
11	97.5	10.2	351	1	MRAY_METH		Q36830 methanobact
12	97	10.1	514	1	COX1_PONPA		P26892 pongo pygma
13	92	9.6	514	1	COX1_PIG		O79876 sus scrofa
14	92	9.6	1060	1	NKCL_MANSE		Q25479 manduca sex
15	91.5	9.5	815	1	NAHL_HUMAN		P19634 homo sapien
16	90.5	9.4	380	1	Y420_METJA		Q37863 methanococc
17	90	9.4	513	1	COX1_HUMAN		P00395 homo sapien
18	89.5	9.3	817	1	NAHL_BOVIN		Q28036 bos taurus
19	89	9.3	370	1	EXOH_RHIME		P33692 rhizobium m
20	89	9.3	383	1	NAPA_ENTHR		P26235 enterococcu
21	88	9.2	385	1	Y421_METTH		Q26521 methanobact
22	88	9.2	426	1	BRNQ_CORGL		O06754 corynebacte
23	87.5	9.1	329	1	SRA7_CABEL		Q93209 caenorhabdi
24	87	9.1	516	1	COX1_BALPH		P24983 balaenopter
25	87	9.1	919	1	PNAL_SCHPO		P09627 schizosacch
26	86.5	9.0	1021	1	TSCC_HUMAN		P55017 homo sapien
27	86	9.0	513	1	COX1_DASNO		O21327 dasypus nov
28	86	9.0	514	1	COX1_SHEEP		O78749 ovis aries
29	86	9.0	516	1	COX1_BALMO		P41293 balaenopter
30	85.5	8.9	274	1	COX1_CHORO		P30671 choristoneu
31	85	8.9	514	1	COX1_BOVIN		P00396 bos taurus
32	84.5	8.8	292	1	MNTC_BACHID		Q9Kd29 bacillus ha
33	84.5	8.8	317	1	HTPX_THEAC		Q9hjv2 thermoplasm

34 84.5 8.8 549 1 COX1_LEITA P14544 leishmania
35 84.5 8.8 816 1 NAHL_RABIT P23791 oryctolagus
36 84 8.8 513 1 COX1_DIDMA P41310 didelphis m
37 84 8.8 513 1 COX1_PAPHA Q92xy2 papio hamad
38 84 8.8 514 1 COX1_EQUAS P92477 equus asinu
39 84 8.8 514 1 COX1_HORSE P50669 equus cabal
40 83.5 8.7 274 1 COX1_CHOFU P40669 choristoneu
41 83.5 8.7 820 1 NAHL_MOUSE O61165 mus musculu
42 83 8.6 502 1 NU2C_MESVI Q9muq6 mesostigma
43 83 8.6 529 1 YABD_SCHPO Q09812 schizosacch
44 83 8.6 1002 1 TSCC_RAT P5018 rattus norv
45 82.5 8.6 217 1 OFCB_BACSU O34878 bacillus su

ALIGNMENTS

RESULT 1
YNES_BACSU
ID YNES_BACSU STANDARD; PRT; 193 AA.
AC Q45064; 1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ynes.
GN YNES.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degree region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter."
RL Microbiology 142:3097-3101(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klein G., Krogsh S., Kumano M.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue G.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nobaç M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takamaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 273234; CAA97604.1; -
 CC EMBL: 299113; CAB13690.1; -
 CC Subtilist: BG11826; yneS.
 CC InterPro: IPR003811; DUF205.
 CC Pfam: PF02660; DUF205; 1.
 CC TIGRFAMs: TIGR00023; DUF205; 1.
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 2 22 POTENTIAL.
 FT TRANSMEM 54 74 POTENTIAL.
 FT TRANSMEM 76 96 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 SQ SEQUENCE 193 AA; 20966 MW; C75803C399B97292 CRC64;
 Query Match 100.08; Score 960; DB 1; Length 193;
 Best Local Similarity 100.08; Pred. No. 1.6e-63;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MLIALLLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNAPRTLGVRAGSVVIAGD 60
 Db 1 MLIALLLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNAPRTLGVRAGSVVIAGD 60
 Oy 61 ILKGTALATLPLMHVDIHPLLAGVAVLGHVFPFPAKFGKAVATSGVLLFYAPLLF 120
 Db 61 ILKGTALATLPLMHVDIHPLLAGVAVLGHVFPFPAKFGKAVATSGVLLFYAPLLF 120
 Oy 121 ITWVAVFFFLYLTKEVSLSSMLTGITYVIYFFVVDYDYLIIWVLLIFVIYRHRANIK 180
 Db 121 ITWVAVFFFLYLTKEVSLSSMLTGITYVIYFFVVDYDYLIIWVLLIFVIYRHRANIK 180
 Oy 181 RIINKTEPKVKWL 193
 Db 181 RIINKTEPKVKWL 193
 RESULT 2
 Y851_STRPN
 ID Y851_STRPN STANDARD; PRT; 213 AA.
 AC Q54916;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein SP0851.
 GN SP0851
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=7785;
 RX MEDLINE=96312346; PubMed=8763932;
 RA Pan X., Fisher M.;
 RT "Cloning and characterization of the parC and parE genes of
 RT Streptococcus pneumoniae encoding DNA topoisomerase IV: role in
 RT fluoroquinolone resistance".
 RL J. Bacteriol. 178:4060-4069(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae".
 RL Science 293:498-506(2001).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 267739; CAA91549.1; -
 CC EMBL: AE007390; AAK74980.1; -
 CC TIGR: SP0851;
 CC InterPro: IPR003811; DUF205.
 CC Pfam: PF02660; DUF205; 1.
 CC TIGRFAMs: TIGR00023; DUF205; 1.
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 2 22 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 164 184 POTENTIAL.
 FT TRANSMEM 168 168 POTENTIAL.
 FT CONFLICT 168 168 S->F (IN REF. 1).
 SQ SEQUENCE 213 AA; 22928 MW; 22CB089C17750818 CRC64;
 Query Match 48.88; Score 468; DB 1; Length 213;
 Best Local Similarity 48.08; Pred. No. 1.1e-27;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 Oy 1 MLIALLLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNAPRTLGVRAGSVVIAGD 60
 Db 1 MITVLLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNAPRTLGVRAGSVVIAGD 60
 Oy 61 ILKGTALATLPLMHVDIHPLLAGVAVLGHVFPFPAKFGKAVATSGVLLFYAPLL 119
 Db 61 FFKGTALATLPLIHFILQGVSPILFGLAVIGTTPFPAKFGKAVATSGVIFGFADIF 120
 Oy 120 FITWVAVFFFLYLTKEVSLSSMLTGITYVI-----YSEFV--HDTYLLIVVLTLLIF 170
 Db 121 CLYLAIFFGALYLGSMISLSSVSTASIAAVIGVLLFPLFGLFSLNYSLSLFAILLALASL 180
 Oy 171 VIYRHRANIKRIINKTEPKVKW 192
 Db 181 IIRHKONARIKKNKTENLVFW 202
 RESULT 3
 YJ73_SYNY3
 ID YJ73_SYNY3 STANDARD; PRT; 222 AA.
 AC P73933;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein sll1973.
 GN SLL1973.
 OS Synecocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90910; BAA17999.1; -;
 CC InterPro: IPR003811; DUF205.
 CC Pfam: PF02660; DUF205; 1.
 CC TIGRfams: TIGR00023; DUF205; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 56 76 POTENTIAL.
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 130 150 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 SQ SEQUENCE 222 AA; 23534 MW; 4718AB2C20833360 CRC64;

 Query Match 36.7%; Score 352.5; DB 1; Length 222;
 Best Local Similarity 42.6%; Pred. No. 3e-19; Indels 15; Gaps 4;
 Matches 87; Conservative 29; Mismatches 73;
 QY 1 MLIALLLIAYLIGSPGLVIGKLAGIDIREHSGNGLGNATNRTLGKVGAGSVVIAGD 60
 DB 6 LILCLLLITYLMGSIPTGLAGKLLGLIDIREHSGKSGTGNVFTLGKPAATAVLAID 65
 QY 61 ILKGLTALPMLHV-DIHPLL-----AGVFAVLGHVPPPIAKFKGKAVATS 108
 DB 66 ISKGMVALVRAIYSGDMLPALPAWONMLTLGVAIAVLGHSHKSIKFKSGGKSVATS 125
 QY 109 GGVLLFAPLFTWVAVFFLYLTKEVSLSSMLTGIYTV--VIYSFFVHDYLLIVVTL 166
 DB 126 LGVFLNINWALGTLATFLVIFTRIVRSLSSIVAAVNGIALALQLPPYLAFTF-L 184
 QY 167 LTFIVYRHRANKRIINKEPKV 190
 DB 185 AGMVIYVRHRTNIERILQCTEPKL 208
 RESULT 4
 Y247_MYCPN STANDARD; PRT; 239 AA.
 AC P75428;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MG247 homolog (H91_orf239).
 GN MPN350 OR MP486.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RC STRAIN=ATCC 29342 / M129;
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelreich R., Hilbert H., Piagens H., Pirkl E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae.";
 RL Nucleic Acids Res. 24:4420-4449(1996).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000047; AAB96134.1; -;
 CC InterPro: IPR003811; DUF205.
 CC Pfam: PF02660; DUF205; 1.
 CC TIGRfams: TIGR00023; DUF205; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 135 155 POTENTIAL.
 FT TRANSMEM 159 179 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 SQ SEQUENCE 239 AA; 27439 MW; 6D4110A8253C9EBB CRC64;

 Query Match 33.2%; Score 318.5; DB 1; Length 239;
 Best Local Similarity 34.2%; Pred. No. 9.4e-17;
 Matches 79; Conservative 45; Mismatches 68; Indels 39; Gaps 6;
 QY 1 MLIALLLIAYLIGSPGLVIGKLAGIDIREHSGNGLGNATNRTLGKVGAGSVVIAGD 60
 DB 9 LLIVFSLVIGYLMGSMFADVFGKILNK-DVRKLGSKNPGATNSIRVFLGKIGLVGLCD 67
 QY 61 ILKSTLTALPFLM-----HVDIH-----PLLAGVFAVLGHVPPPIAKFKGKAV 105
 DB 68 ALKGFALFVFSFLFSFWLQOYLVNVOYKVVYLYLSCFAATIGHIPFLYFKFGKRAI 127
 QY 106 ATSGVLLFAPLFTWVAVFFLYLTKEVSLSSMLTGIYTVI-----YSFFV-- 155
 DB 128 ATTSGLLALISLWFWFVCLVLLVLTITKYVSLASLVTFILAILIIVPWLVDLYFFKP 187
 QY 156 -----HDTYLLIVVTL-----LTFIVYRHRANKRIINKEPKVKWL 193
 DB 188 NPINAISYQNDWYIILFVFLWYPLTIAVFWLHRKNIHRLNKNENKVTQL 238
 RESULT 5
 Y676_AQUAE STANDARD; PRT; 192 AA.
 AC O66905;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AQ_676.
 GN AQ_676.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 OC Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RC MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus".
 RL Nature 392:353-358(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AE000701; AAC06869.1; --
 DR InterPro; IPR003811; DUF205.
 DR Pfam; PF02660; DUF205; 1.
 DR TIGRfams; TIGR00023; DUF205; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 80 100 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 SQ SEQUENCE 192 AA; 20940 MW; EAD53C4016D63C00 CRC64;

Query Match 31.5%; Score 302; DB 1; Length 192;
 Best Local Similarity 37.7%; Pred. No. 1.2e-15;
 Matches 75; Conservative 36; Mismatches 66; Indels 22; Gaps 6;

Qy 1 MLIALIILAVLIGSLIPGLVGLKAKGIDIREHSGNIGATNAPRTLGVRAGSVVIAGD 60
 Db 1 MKALELVIFAYLLGSITGEVIAKL-KGVDLRNVGSGNIGATNTRALCKKYGVLVFLD 59
 Qy 61 ILKGLTALPFLMHVDIHL-----LAGVAVLGHVFPFIFAKFGKAVATSGGVLLF 114
 Db 60 FLKG-----FIPALIAVKSFGDSWVLTFTGLASVLGHMYPVFPFGKGGKAVATGALGVVFA 115
 Qy 115 YAPLFIYVAVFFIFLVL-----TKFVLSLSMLTGIVTVISFV-HDTVLLIVVTLT 168
 Db 116 VSP-----SVALFSLVGLIFLWKRYVSLASITATISAFLEFLFVAGVPVNVFLMAIVIG 170
 Qy 169 IFVIYHRHRANIKRIINKE 187
 Db 171 ALIIYHRENINRLTGRE 189

RESULT 6
 Y247_MYCGE
 ID Y247_MYCGE STANDARD; PRT; 239 AA.
 AC P47489;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MG247.
 GN MG247.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2057;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Kirsch E.F.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium."
 RL Science 270:397-403(1995).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U39703; AAC71467.1; --
 DR TIGR; MG247; --
 DR InterPro; IPR003811; DUF205.
 DR Pfam; PF02660; DUF205; 1.
 DR TIGRfams; TIGR00023; DUF205; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 135 155 POTENTIAL.
 FT TRANSMEM 159 179 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 SQ SEQUENCE 239 AA; 27489 MW; D78CE976DEF621FD CRC64;

Query Match 29.6%; Score 284.5; DB 1; Length 239;
 Best Local Similarity 32.2%; Pred. No. 2.8e-14;
 Matches 75; Conservative 43; Mismatches 72; Indels 43; Gaps 6;

Qy 3 IALLIILA-----YLIGSLIPGLVGLKAKGIDIREHSGNIGATNAPRTLGVRAGSVVIA 58
 Db 7 IAILVIFSLASGLLGSITFIADIFSKLKK-NVREFGSKNPGATNSMRVFLKIGFLVAI 65
 Qy 59 GDILKGLTALPFLM-----HVDIHPLLAGVFAVLGHVFPFIFAKFGCK 103
 Db 66 FDAFKGFAFLLTWLFRGLQGLYTEKVYOSTYFLSYLSCFAATIGHIPLYPFKFGKK 125
 Qy 104 AVATSGGVLLFYAPLLFTVAVFFIFLYLTKFVLSLSMLTGIVTVI-----YSFFV 155
 Db 126 AIATGSGSLAISLWFLICLLIWMITLTKYVSLASLITFFVLAVILLIPWLDLYLFF 185
 Qy 156 -----HDYLLIVTVL-----LTIFVIYHRHRANIKRIINKEPKVKWL 193
 Db 186 NSDPLKSLITYQENYIILFCLWYVPLTVVFWLHRLHRIILHKGESKITQL 238

RESULT 7
 YGIH_HAEIN
 ID YGIH_HAEIN STANDARD; PRT; 199 AA.
 AC P44603;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein HI0266.
 GN HI0266.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=RJ / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirsch E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RD.";
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY. STRONG, TO E.COLI YGIH.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Matches 51: Conservative 34; Mismatches 79; Indels 44; Gaps 10;

OY 10 AYLGISPSGLIVKGLAKGIDIREHSGN-----LGNATNAPRLGVKAGSVVIAGD 60
 Db 308 ATMIATPTGVKFSWLAFL-----HGSNKTWSAAILMALGFIFLF-TVGGLTG-IVLANS 361

OY 61 ILKGTATATLPLMHVDIPLLAGVFAVLG---HVPEIFAKFGKAVATSGGVLLFYAP 117
 Db 362 SLDIVLHDTYVVVAHFVLSMGAVFALMGGFIIHWFPLFSGYTLDTQ-----YAK 411

OY 118 LLFTMTAVVFIFLYLT-----KFVSLSSM-----LGTIVTVIYFFVHDYTLIVVTL 167
 Db 412 IHFTM-----PIGNLTFPFOHFLGSLGMPRRYSYDPDAYTTWNILSAGSFISLTAVML 467

OY 168 TIFVYHRANIKRI--INKTEPKVKWL 193
 Db 468 MIFMIWEAFASKRVPMPVOPSTSLLEWL 495

RESULT 13

COX1_PIG * STANDARD; PRT; 514 AA.

AC 079876; Q9TOR4;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN MT001 OR COI.
 OS Sus scrofa (Pig).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98404150; PubMed=9732457;
 RA Ursing B.M., Arnason U.
 RT "The complete mitochondrial DNA sequence of the pig (Sus scrofa).";
 RL J. Mol. Evol. 47:302-306(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Landrace; TISSUE=Heart;
 RX MEDLINE=99365306; PubMed=10433971;
 RA Lin C.S., Sun Y.L., Liu C.Y., Yang P.C., Chang L.C., Cheng I.C.,
 MAO S.J.T., Huang M.C.;
 RT "Complete nucleotide sequence of pig (Sus scrofa) mitochondrial genome
 and dating evolutionary divergence within artiodactyla.";
 RL Gene 236:107-114(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Meishan, and Swedish wild boar;
 RA Kijas J.M.H., Andersson L.;
 RT "A phylogenetic study of the origin of the domestic pig estimated from
 the near complete mtDNA genome.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferrocyclochrome
 c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AJ002189; CA05231.1; -
 DR EMBL; AF034253; A034187.1; -
 DR EMBL; AF034203; A034218.1; -
 DR EMBL; AF034200; A034217.1; -
 DR HSPSP; P00396; Z0CC.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 61 61 IRON (HEME A) (PROBABLE).
 FT METAL 240 240 COPPER B (PROBABLE).
 FT METAL 244 244 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 291 291 COPPER B (PROBABLE).
 FT METAL 376 376 IRON (HEME A3) (PROBABLE).
 FT METAL 378 378 IRON (HEME A) (PROBABLE).
 FT CONFLICT 49 50 GD -> PH (IN REF. 1).
 SQ SEQUENCE 514 AA; 56958 MW; 6B5008565248CF3A CRC64;

Query Match 9.8%; Score 92; DB 1; Length 514;
 Best Local Similarity 22.1%; Pred. No. 5.4;
 Matches 45; Conservative 38; Mismatches 85; Indels 36; Gaps 8;

OY 10 AYLGISPSGLIVKGLAKGIDIREHSGNIGATNAFR-----TLGVKAGSVVIAGD 60
 Db 308 ATMIATPTGVKFSWLAFL-----HG-GNIKWSPAMLMALGFLFTVGGLTG-IVLANS 361

OY 61 ILKGTATATLPLMHVDIPLLAGVFAVLG---HVPEIFAKFGKAVATSGGVLLFYAP 117
 Db 362 SLDIVLHDTYVVVAHFVLSMGAVFALMGGFVHWFPLFSGYTLNQ-----WAK 411

OY 118 LLFTMTAVVFIFLYLTKFVLSLSM-----LGTIVTVIYFFVHDYTLIVVTLTIFV 171
 Db 412 IHFVIMFVGNVMTFFPOHFLGSLGMPRRYSYDPDAYTANNTISSMGSFISLTAVMLIFI 471

OY 172 IYRHRANIKRI--INKTEPKVKWL 193
 Db 472 IWEAFASKREVSVELTSTNLEWL 495

RESULT 14

NKCL_MANSE STANDARD; PRT; 1060 AA.

ID NKCL_MANSE
 AC Q25479;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Bumetanide-sensitive sodium-(Potassium)-chloride cotransporter
 DE (NA-K-CL symporter).
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
 OC Ditrysia; Sphingioidea; Sphingidae; Manduca;
 OX NCBI_TaxID=7130;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Malpighian tubules;
 RX MEDLINE=95035837; PubMed=7550244;
 RA Reagan J.D.;
 RT "Molecular cloning of a putative Na(+)-K(+)-2Cl-cotransporter from
 the Malpighian tubules of the tobacco hornworm, Manduca sexta.";
 RL Insect Biochem. Mol. Biol. 25:875-880(1995).
 CC -1- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
 MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
 IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U17344; AAA75600.1; --
 DR InterPro: IPR002293; AA/rel_pmaseel.
 DR InterPro: IPR004842; KCL_cotransport.
 DR InterPro: IPR002443; NaKCl_transporter.
 DR PRINTS: PR01207; NAKCLTRNSPRT.
 DR TIGRFAMs: TIGR00930; 2a30; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT DOMAIN 1 122 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 123 133 POTENTIAL.
 FT DOMAIN 144 153 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 154 174 POTENTIAL.
 FT DOMAIN 175 197 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 198 218 POTENTIAL.
 FT DOMAIN 219 249 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 250 270 POTENTIAL.
 FT DOMAIN 271 275 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 276 296 POTENTIAL.
 FT DOMAIN 297 331 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 332 352 POTENTIAL.
 FT DOMAIN 353 367 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 368 388 POTENTIAL.
 FT DOMAIN 389 431 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 432 452 POTENTIAL.
 FT DOMAIN 453 497 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 498 518 POTENTIAL.
 FT DOMAIN 519 562 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 563 583 POTENTIAL.
 FT DOMAIN 584 642 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 643 663 POTENTIAL.
 FT DOMAIN 664 881 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 882 902 POTENTIAL.
 FT DOMAIN 903 1060 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1060 AA; 117392 MW; 4C39D1F0B645FFEF CRC64;
 Query Match 9.6%; Score 92; DB 1; Length 1060;
 Best Local Similarity 23.1%; Pred. No. 9.8;
 Matches 57; Conservative 34; Mismatches 84; Indels 72; Gaps 12;
 QY 2 LIALLI-ILAYLI-----GSI--PSGLIVGKLAKGIDIREHSGNIGATNAFTLG-- 49
 DB 368 LLALLISMSYTLMLVLFAGGGALDRASGNITDLIVNGVTVDYSSVLCALNNTCEYGLH 427
 QY 50 -----VKAGSVIAGDILAGTLATLPFLMHVDIHPLLAGVFAVLGHVEP---IPAK 98
 DB 428 NSYSVMQLMSANGPFYFGCGWATLSTALTNLVS---PRLIQALGV-DRITPGLIFFRS 483
 QY 99 FKGGKAVATSGGVLLFYAPLLFI-----TWAVFFI----- 129
 DB 484 PYGRHCEPYRGVYLTFFVSLFLLIADNTIAPLISNEYSVALSNFCTFHRALVRPLG 543
 QY 130 -----FLYLKFLVSLNSLGIYTVYFFVHDYTLIVTLTIFVIYHRANIKRIINK 185
 DB 544 WRPTFYRNWMLSLAGL---MCVAILMLVH--WNVSLVTFAIFTLY-----LIVHY 591
 QY 186 TEPKVKW 192
 DB 592 RRPDVNW 598
 RESULT 15

NAHL_HUMAN STANDARD; PRT; 815 AA.
 ID NAHL_HUMAN AC P19634;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sodium/hydrogen exchanger 1 (Na⁺)/H⁺ exchanger 1 (NHE-1) (Na⁺/H⁺ antipporter, amiloride-sensitive) (APNH).
 GN SLC9A1 OR NHE1 OR APNH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=89106219; PubMed=2536298;
 RA Sardet C., Franchi A., Pouyssegur J.;
 RT "Molecular cloning, primary structure, and expression of the human growth factor-activatable Na⁺/H⁺ antipporter.";
 RL Cell 56:271-280(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=90140739; PubMed=2154036;
 RA Sardet C., Counillon L., Franchi A., Pouyssegur J.;
 RT "Growth factors induce phosphorylation of the Na⁺/H⁺ antipporter, glycoprotein of 110 kD.";
 RL Science 247:723-726(1990).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=91293066; PubMed=1712287;
 RA Tse C.-N., Ma A.I., Yang V.W., Watson A.J.M., Levine S., Montrose M.H., Potter J., Sardet C., Pouyssegur J., Donowitz M.;
 RT "Molecular cloning and expression of a cDNA encoding the rabbit ileal villus cell basolateral membrane Na⁺/H⁺ exchanger.";
 RL EMBO J. 10:1957-1967(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=94111706; PubMed=8283968;
 RA Fliegel L., Dyck J.R., Wang H., Fong C., Haworth R.S.;
 RT "Cloning and analysis of the human myocardial Na⁺/H⁺ exchanger.";
 RL Mol. Cell. Biochem. 125:137-143(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20375279; PubMed=10913675;
 RA Garden O.A., Musk P., Worthington-White D.A., Dewey M.J., Rich I.N.;
 RT "Silent polymorphisms within the coding region of human sodium/hydrogen exchanger isoform-1 cDNA in peripheral blood mononuclear cells of leukemia patients: A comparison with healthy controls.";
 RL Cancer Genet. Cytogenet. 120:37-43(2000).
 CC -!- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED BY ACTIVE METABOLISM OR TO COUNTER ADVERSE ENVIRONMENTAL CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL TRANSDUCTION.
 CC -!- SUBUNIT: Interacts with tescalcin.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: KIDNEY AND INTESTINE.
 CC -!- PTM: PHOSPHORYLATED (POSSIBLE).
 CC -!- MISCELLANEOUS: INHIBITED BY AMILORIDE AND 5-AMINO-SUBSTITUTED DERIVATIVES AND ACTIVATED IN A COOPERATIVE FASHION BY INTRACELLULAR H⁺. FULLY ACTIVE AT ACIDIC PH, THE ANTIPOPPER IS VIRTUALLY TURNED OFF AT NEUTRAL PH. IN QUIESCENT CELLS UPON GROWTH FACTOR STIMULATION, THE APPARENT AFFINITY FOR INTERNAL H⁺ IS INCREASED, RESULTING IN A PERSISTENT RISE IN CYTOPLASMIC PH.
 CC -!- SIMILARITY: BELONGS TO THE NA⁺/H⁺ EXCHANGER FAMILY.
 CC -!- CAUTION: THE NUMBER, LOCALIZATION AND DENOMINATION OF HYDROPHOBIC DOMAINS IN THE NA⁺/H⁺ EXCHANGERS VARY AMONG AUTHORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M81768; AAB59460.1; ALT_SEQ.
DR EMBL; S68616; AAC60606.1; -
DR EMBL; AF141350; AAF21350.1; -
DR EMBL; AF141351; AAF21351.1; -
DR EMBL; AF141352; AAF21352.1; -
DR EMBL; AF141353; AAF21353.1; -
DR EMBL; AF141354; AAF21354.1; -
DR EMBL; AF141355; AAF21355.1; -
DR EMBL; AF141356; AAF21356.1; -
DR EMBL; AF141357; AAF21357.1; -
DR EMBL; AF141358; AAF21358.1; -
DR EMBL; AF141359; AAF21359.1; -
DR EMBL; AF146430; AAF25592.1; -
DR EMBL; AF146431; AAF25593.1; -
DR EMBL; AF146432; AAF25594.1; -
DR EMBL; AF146433; AAF25595.1; -
DR EMBL; AF146434; AAF25596.1; -
DR EMBL; AF146435; AAF25597.1; -
DR EMBL; AF146436; AAF25598.1; -
DR EMBL; AF146437; AAF25599.1; -
DR EMBL; AF146438; AAF25600.1; -
DR EMBL; AF146439; AAF25601.1; -
DR PIR; A31311; A31311.
DR Genew; HGNC:11071; SLC9A1.
DR MIN; 107310; -
DR InterPro; IPR000676; NaH_Exchng.
DR InterPro; IPR004709; NaH_exchang3.
DR Pfam; PF00999; Na_H_Exchange; 1.
DR PRINTS; PR01084; NAHEXCHNGR.
DR TIGRFAMs; TIGR00840; b_cpai.1.
KW Transmembrane; Glycoprotein; Transport; Antiport; Sodium transport;
KW Multigene family; Phosphorylation.

FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 35 M1 (POTENTIAL).
FT DOMAIN 36 107 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 108 127 M2 (POTENTIAL).
FT DOMAIN 128 129 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 130 149 M3 (POTENTIAL).
FT DOMAIN 150 154 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 155 174 M4 (POTENTIAL).
FT DOMAIN 175 191 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 192 211 M5 (POTENTIAL).
FT DOMAIN 212 227 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 228 247 M5A (POTENTIAL).
FT DOMAIN 248 256 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 257 276 M5B (POTENTIAL).
FT DOMAIN 277 294 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 295 315 M6 (POTENTIAL).
FT DOMAIN 316 338 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 339 358 M7 (POTENTIAL).
FT DOMAIN 359 386 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 387 406 M8 (POTENTIAL).
FT DOMAIN 407 410 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 411 430 M9 (POTENTIAL).
FT DOMAIN 431 480 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 481 500 M10 (POTENTIAL).
FT DOMAIN 501 815 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 815 AA; 90763 MW; 02EC748C79DF6526 CRC64;

Query Match 9.58; Score 91.5; DB 1; Length 815;
Best Local Similarity 20.78; Pred. No. 8.6;
Matches 58; Conservative 50; Mismatches 79; Indels 93; Gaps 12;
QY 3 IALLIILA-----YLIGSIPS-----GLIVGNLAKGID----- 30

Db 105 ISLWILLACLMKIGFHVPTISSIVPESCLLIIVVGLLVGLIKGVGETPPPLQSDVFFLF 164
Qy 31 -----IREHG-----SGNLGATNAPRTLGVRAGSVVIAG----- 59
Db 165 LLPPIILDAGYFLPLROPTENLGTILIFAVVGTLWNAFFLGLMYAVCLVGEQINNI 224
Qy 60 --DILKGTALATLPFL-----MHVD--IHPLLAG-----VFVGLGHVFIKPK 100
Db 225 LDNLFGSIISAVDPVAVLAVFEEIHNELHLVFGESLLNDAVTVVLYHLFEFANYE 284
Qy 101 GSKAVATSGGVLLFYAPLLFITMVAVFP--IFLYLTKEVSLSSMLTGYIYTYFFVHDT 158
Db 285 HVGIVDFLGSFSEFFVALGVGVVGYVIAAFTSRFTSHRVEIPLFVFLYSTMAYLS 344
Qy 159 ----YLLIVVTLLTIFVIYRH--RANIKRIINKTEPKVKW 192
Db 345 AELFHLSCIMALIASGVVMRPYVEANIS---HKSHTTIKY 381

Search completed: December 26, 2002, 00:54:17
Job time : 140.382 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 00:00:26 : Search time 623.209 Seconds
(without alignments)
63.810 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLTALLIILAYLIGSIPSGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rviro:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568.5	59.2	198	16 Q92C68	Q92C68 listeria in
2	566.5	59.0	198	16 Q8Y7J3	Q8Y7J3 listeria mo
3	464.5	48.4	202	16 Q99UC5	Q99UC5 staphylococ
4	453	47.2	213	16 Q9A070	Q9A070 streptococ
5	445	46.4	213	16 Q9CGW4	Q9CGW4 lactococcc
6	430	44.8	194	16 Q8RFY9	Q8RFY9 fusobacteri
7	385.5	40.2	168	2 Q9X972	Q9X972 streptococ
8	373.5	38.9	198	16 Q8R9J2	Q8R9J2 thermoaer
9	339.5	35.4	226	16 Q8YZG8	Q8YZG8 anabaena sp
10	333	34.7	205	16 Q8UFU1	Q8UFU1 agrobacteri
11	322	33.5	206	16 Q9KCD3	Q9KCD3 bacillus ha
12	309.5	32.2	196	16 Q9XIF9	Q9XIF9 thermotoga
13	301.5	31.4	199	2 Q9ZAF3	Q9ZAF3 thermus the
14	301.5	31.4	203	16 Q9ZQL7	Q9ZQL7 rhizobium m
15	301	31.4	201	16 Q8YC64	Q8YC64 brucella me
16	299	31.1	208	16 Q9KUJ7	Q9KUJ7 vibrio chol

17	296	30.8	198	16 Q9RSV1	Q9RSV1 deinococcus
18	292.5	30.5	195	16 Q98M84	Q98M84 rhizobium l
19	286	29.8	216	16 Q8Z167	Q8Z167 yersinia pe
20	284.5	29.6	218	16 Q9A5K1	Q9A5K1 caulobacter
21	271.5	28.3	198	16 Q97G69	Q97G69 clostridium
22	270.5	28.2	257	16 Q9PQ85	Q9PQ85 ureaplasma
23	270	28.1	207	16 Q8XWC8	Q8XWC8 ralstonia s
24	266.5	27.8	201	16 Q9CKC7	Q9CKC7 pasteurella
25	263.5	27.4	202	16 Q9FIE4	Q9FIE4 campylobact
26	261.5	27.2	224	16 Q98Q6	Q98Q6 mycoplasma
27	257.5	26.8	203	16 Q8XGX7	Q8XGX7 salmonella
28	257	26.8	200	16 Q9JZG9	Q9JZG9 neisseria m
29	257	26.8	200	16 Q9JUL4	Q9JUL4 neisseria m
30	252	26.2	189	16 Q915V6	Q915V6 pseudomonas
31	198.5	20.7	203	16 Q9X109	Q9X109 thermotoga
32	165	17.2	195	16 Q9RS57	Q9RS57 deinococcus
33	141	14.7	559	16 Q9RW79	Q9RW79 deinococcus
34	133	13.9	425	17 Q8TH23	Q8TH23 methanosarc
35	115	12.0	463	16 Q34474	Q34474 bacillus su
36	110.5	11.5	421	16 Q9X0D5	Q9X0D5 thermotoga
37	108.5	11.3	243	17 Q8TQJ0	Q8TQJ0 methanosarc
38	104.5	10.9	963	16 Q8YU47	Q8YU47 anabaena sp
39	104	10.8	325	17 Q973Q6	Q973Q6 sulfolobus
40	101.5	10.6	381	17 Q9H1Z1	Q9H1Z1 thermoplasma
41	103	10.4	374	17 Q97X02	Q97X02 sulfolobus
42	103	10.4	671	17 Q8TJA8	Q8TJA8 methanosarc
43	99	10.3	477	16 Q8XP64	Q8XP64 clostridium
44	99	10.3	513	8 Q9TX97	Q9TX97 pongo pygma
45	98.5	10.3	276	17 Q9UZ69	Q9UZ69 pyrococcus

ALIGNMENTS

RESULT 1
Q92C68 ID Q92C68 PRELIMINARY; PRT; 198 AA.
AC Q92C68; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein lin1323.
GN LIN1323.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapatk.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Medjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
Vazquez-Boland J.-A., Voss H., Wehland J.J., Cossart P.;
Comparative genomics of Listeria species.;
RL Science 294:849-852(2001).
EMBL; AL596168; CAC96554.1;
L1323; LIN01323;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRfams: TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21632 MW; B161D1055B203406 CRC64;

Query Match 59.2%; Score 568.5; DB 16; Length 198;

Best Local Similarity 57.7%; pred. No. 2.3e-40;
Matches 112; Conservative 35; Mismatches 44; Indels 3; Gaps 1;

	Oy	3	IALLIILAYLIGSIPSGLIIVGKLAKGIDIREHSGSNLGATNPRTLGVKAGSVIADGL	62
			: : : : : : : : : : : : :	
	Db	5	LILLSLAYVIGSIPSGLTWGIIFYKKDIRDFGSNGLGNATNSFVLGVKAGSIVTMDIL	64
	Oy	63	KGTLLATALPFLMHVDTHP --- LLAGVFVAVLGHVPDIFAFKFGKGKAVATSGGVLLFPAPLL	119
			: : :: : : : : : : : : :	
	Db	65	KGTVAITLLPPFFOLFQNLNNHHFWLLTGAFATIGHSFPLFAGFRGGKAVATSAGVILAYAPLL	124
	Oy	120	FITMWAVPFILYLTKEVSLSSMLTCIYVIYSFFVHDHYLLIVVTLLTFIVYIHRRAM	179
			: - : : : : : : : : : : : : :	
	Db	125	FVALVVFTLKISRYVLSMMICALAALITISFFMGDMILLIIVACIALFIWFWRHRAM	184

Qy	180 KRIINKTEPKVKWL	193
	: :	
Db	185 TRIRNGEEPKIKWM	198

RESULT 2	
Q8Y7J3	
ID	Q8Y7J3
AC	Q8Y7J3;
DT	01-MAR-2002 (PRELIMINARY; PRT; 198 AA.
DT	01-MAR-2002 (TRENBLrel. 20, Created)
DT	01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE	Hypothetical protein Imol284.

GN Listeria monocytogenes.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacteriacea; Firmicutes; Bacillus/Clostridium group; Bacillales;
OX Listeriaceae; Listeria.
NCBI_TaxID=1639; Listeria

RP	SEQUENCE FROM N.A.
RC	STRAIN-EGD-E / SEROVAR 1/2A;
RC	MEDLINE-21537279; PubMed-1679669;
RX	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA	Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA	Charbit A., Chetoui F., Couve E., de Daruvar A., Deboux P.,
RA	Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA	Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA	Jones L.-M., Kaerst U., Klett J., Kuhn M., Kunst F., Kurupkat G.,
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA	Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT	"Comparative genomics of <i>Listeria</i> species.";
RL	Science 294:849-852(2001).
DR	EMBL; AL591978; CAC99362.1; -.
DR	ListiList; LMO01284; -.
DR	InterPro: IPR003811; DUF205.
DR	Pfam: PF02660; DUF205; 1.
DR	TIGRFAMS; TIGR00023; DUF205; 1.
DR	HyPOthetical protein; Complete proteome.
SW	SEQUENCE 198 AA; 21599 MW; 7807B5406DF05GD1 CRC64;
KQ	

Query Match	59.0%	Score 566.5;	DB 16;	Length 198;
Best Local Similarity	58.8%	Pred. No. 3.4e-40;		
Matches 114;	Conservative 33;	Mismatches 44;	Indels 3;	Gaps 2;

Qy	3	TALLIILAYLIGSPISGLIVGKIAKAGIDIREHSGSNLGATNAFRTLGVKAGSVIAGDIL	62
Db	5	LILLSLAYLIGSPISGLWICKIFRYKKDIRFEGSGNLGATNSPVLIGKAGSVITVMDIL	64
Qy	63	KGTLATLALPFL--MHVDIH-PLIAGVPFVLGHVFPPIAKFKGKAVATSGGVLLFYAPLL	119
Db	65	KGTVATLLPFFQLNVDRHFWLLTGAFATIGHSPFLPAGFRGKAVATSGVILAYAPLL	124
Qy	120	PITWAVVFELFLYLTAKFVSUSSMLTGLTYTVIYSEFFVHDYTLIIWVLLTTFVIVRHRANI	179
Db	125	FVAALVFLVTLKLSKVYSUSSMIGALAAALISLEFGDWTLIIVLVACIALFVIVRHRANI	184

```
Qy 180 KRIINKTEPKVKWL 193
    ||| |||:||:
Db 185 TRIRNGEEPKIKWM 198
```

RESULT 3				
Q99UC5	PRELIMINARY;	PRT;	202 AA.	
ID	Q99UC5			
AC	Q99UC5;			
DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE	Hypothetical protein SAV1353.			
GN	SAV1353 OR SAIL87.			
OS	Staphylococcus aureus (strain Mu50 / ATCC 700699), and			
OS	Staphylococcus aureus (strain N315).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC	Staphylococcus.			
OX	NCBI_TaxID:158878; 158879;			

RP SEQUENCE FROM N.A.
RC SPECIES: *S.aureus* (strain Mu50), and *S.aureus* (strain N315);
RX MEDLINE#21311952; PubMed#11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iitan J.-Q., Ito T.,
Rakazutani M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Kanamori M., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "whole genome sequencing of methicillin-resistant *Staphylococcus aureus*,"
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003362; BAB57515.1; -.
DR EMBL: AP003133; BAB42445.1; -.
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TRIGRAMS: TIGR00023; DUF205; 1.
DR Hypothetical protein; Complete
KW SEQUENCE 202 AA; 22332 MW; A9DA126B5731749C CRC64;

Query Match	48.4%;	Score 464.5;	DB 16;	Length 202;
Best Local Similarity	46.5%;	Pred. NO. 1.3e-31;		
Matches 94;	Conservative 40;	Mismatches 59;	Indels 9;	Gaps 2;

Qy	1	MLIALLLIAYLIGSIPSGLIIVGKLAKGIDIRHSGSNGIATNAPFTLGVKAGSVVIAGD	60
Db	1	MMHIVMLLSYLIGAFPSGVFIKLEFPKDIROFGSGNTGATNSFVLGAPAGFLVTPJD	60
Qy	61	ILKGTFLATALPFLM-----IHPLLAGVFAVLGHVPPIFAKFKGGKAVATSGGV	111
Db	61	IFKGFIVFPFLWLPVWADGPISFTFTNGLIVGLFALGHVLPVYLPKGGKAVATSGV	120
Qy	112	LLFVAPLLFTIMVAVFFIFLYLTKFVLSMLGTYTVIYSPFVHDTYLLIVVLLTTFV	171
Db	121	VLG/NPILLILATIFEIFVLKFKYVSLASIVAAICCVIGSLIQDYILLVVSFLYSIIL	180
Qy	172	IYHRANIKRIINKTEPKVWL	193
Db	181	IIRHSNIARIFRGEPEKIKWM	202

RESULT 4	
ID	Q9A070
AC	Q9A070;
DT	01-JUN-2001 (TRENBLrel. 17, Created)
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE	Hypothetical protein SPY0908.
GN	SPY0908.
OS	Streptococcus pyogenes

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Perreux C., Sezate S., Smorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006539; AAK33824.1; -;
 DR InterPro: IPR003811; DUF205.
 DR TIGRFAMs: TIGR00023; DUF205; 1.
 DR Pfam: PF02660; DUF205.
 DR Hypothetical protein; Complete proteome.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 213 AA; 23369 MW; 6A9881232A09766A CRC64;

Query Match 47.2%; Score 453; DB 16; Length 213;
 Best Local Similarity 47.0%; Pred. No. 1.3e-30;
 Matches 95; Conservative 34; Mismatches 63; Indels 10; Gaps 3;
 OY 1 MLIALLIILAYLIGSIPGLVGLKAGIDIREHSGNGLGNATNAPRTLGVKAGSVVIAGD 60
 DB 1 MKLLFIYIALLGSIPTGLMIGQFYHINLREHSGNGTGTNTFRILGVKAGTATLAI 60
 OY 61 ILKGTALATLAPFLMHV-DIHPLLAGVAVLGHVFFIFAKFKGKAVATSGVLLFYAPLL 119
 DB 61 MKFGTSLILPIFGMTSSISSTAIGFAVLTGTFIFANFKGKAVATSGVLLGFAPLY 120
 OY 120 FITWAVFEFLYLTKEFVLSMLTGIVTVI-----YSFFVHDYLLIVVTLTIF 170
 DB 121 LFFLASIFVLVLFMSLSVSWAIVGLSVLTFPAIHFLPNYDFLTIVILLAFI 180
 OY 171 VIYHRANIKRIINTEPKVK 192
 DB 181 IIRHNDNISRIKHTEINLPW 202

RESULT 5
 O9CGW4 PRELIMINARY; PRT; 213 AA.
 AC O9CGW4:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein ykac.
 GN YKAC OR L40978.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauger S., Jallion O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 lactis ssp. lactis IL1403.";
 RL Genome Res. 11:731-753(2001).
 DR EMBL: AE006332; AAK05076.1; -;
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 DR TIGRFAMs: TIGR00023; DUF205; 1.
 DR Hypothetical protein; Complete proteome.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 213 AA; 23362 MW; 967258F00F31AE3 CRC64;

Query Match 46.4%; Score 445; DB 16; Length 213;
 Best Local Similarity 45.5%; Pred. No. 6.1e-30;
 Matches 91; Conservative 39; Mismatches 60; Indels 10; Gaps 2;

OY 1 MLIALLIILAYLIGSIPGLVGLKAGIDIREHSGNGLGNATNAPRTLGVKAGSVVIAGD 60
 DB 2 LTIILLIASYLLGAIPGLWIGKIFFKKNLHDYSGNGTGTNTFRILGVKAGISVFAFD 61
 OY 61 ILKGTALATLAPFLMHV-DIHPLLAGVAVLGHVFFIFAKFKGKAVATSGVLLFYAPLL 119
 DB 62 LLKGTALATLAPFLPHINGVSPILFGLLAVIGHTTFIDRFKGGKAVATSGVILGFSPLF 121
 OY 120 FITWAVFEFLYLTKEFVLSMLTGIVTVI-----YSFFVHDYLLIVVTLTIF 170
 DB 122 LIYLVFIIIVLWLFMSLSLSVGVAVFALLGILIFPSIGFILTSYDLFSIIIFVLAII 181
 OY 171 VIYHRANIKRIINTEPKVK 190
 DB 182 IILRHTNLKRIKNHCSLV 201
 RESULT 6
 O8RFY9 PRELIMINARY; PRT; 194 AA.
 AC O8RFY9:
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical membrane-spanning protein FN0537.
 GN FN0537.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteriia; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fongstein M., Kyrpides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010565; AAL94733.1; -;
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 194 AA; 21343 MW; DE327E0AE835915 CRC64;

Query Match 44.8%; Score 430; DB 16; Length 194;
 Best Local Similarity 44.4%; Pred. No. 1e-28;
 Matches 87; Conservative 41; Mismatches 62; Indels 6; Gaps 4;

OY 1 MLIALLIILAYLIGSIPGLVGLKAGIDIREHSGNGLGNATNAPRTLGVKAGSVVIAGD 60
 DB 1 MAFFCFIVLTYPFGAIPSGVWIGKAFKGVVDVDRYDGSKNAGTNSYRVLGAKLGAVVLMD 60
 OY 61 ILKG--TLATLAPF-LMHVDIHPLLAGVAVLGHVFFIFAKFKGKAVATSGVLLFYAP 117
 DB 61 VLKGFIPLYIASKFNLYNDL--VILGLVAILAHTFSCFISFKGKGVATSGVFLFLIP 118
 OY 118 LFIWVAVFFFLYTLTFVSVLSMLTGIVTVIYFFVH-DYLLIVVTLTIFVYRHR 176
 DB 119 VITEILLAIFILVAVFTKYVSLASITAAPLLPIFTFFTHKDSYLSALSIVIAFVYRHK 178
 OY 177 ANIKRIINTEPKVKW 192
 DB 179 TNISRLSLGTENKFKF 194

RESULT 7
 Q9X972 PRELIMINARY; PRT; 168 AA.
 ID Q9X972
 AC Q9X972:
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

```

DB 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Hypothetical 17.9 kDa protein (Fragment).
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH1.
RA Vriesema A.J., Dankert J., Zaai S.A.;
RT "Isolation and characterization of promoter regions from Streptococcus
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ236899; CAB40549.1;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
KW Hypothetical protein.
FT NON_TER 168
SQ SEQUENCE 168 AA; 17921 MW; A07262BD799A478A CRC64;

Query Match 40.2%; Score 385.5; DB 2; Length 168;
Best Local Similarity 51.7%; Pred. No. 4.9e-25;
Matches 75; Conservative 30; Mismatches 39; Indels 1; Gaps 1;

QY 7 ILAYLIGSPGLVIGKLAGKIDIREHSGNGLGATNAPRTLGVKAGSVVVIAGDILKGTLL 66
DB 8 LILAYLIGSPGLVIGKLAGKIDIREHSGNGLGATNAPRTLGVKAGSVVVIAGDILKGTLL 67
QY 67 ATALPFLMVD-IHPLLAGVFAVLGHVFFIFAKFGKGAATSGGVLLFYAPLFFITMVA 125
DB 68 ATLLPFLHNGISPMIFGLAVLGHVFFIFAKFGKGAATSGGVLLFYAPLFFITMVA 127
QY 126 VFIFLYLTKFVSLSSMLTGITVI 150
DB 128 IFIVLYLGSMSISIVVAGFAII 152

RESULT 8
QY 8 Q89J2 PRELIMINARY; PRT; 198 AA.
AC Q89J2
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Hypothetical protein TFE1618.
CN TFE1618.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB47 / JCM11007;
RX MEDLINE-21992816; PubMed-11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL: AE013117; AAM24820.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21016 MW; 252C3FF7D512BF02 CRC64;

Query Match 38.9%; Score 373.5; DB 16; Length 198;
Best Local Similarity 43.5%; Pred. No. 5.9e-24;
Matches 83; Conservative 33; Mismatches 74; Indels 1; Gaps 1;

QY 1 MLIALIILAYLIGSPGLVIGKLAGKIDIREHSGNGLGATNAPRTLGVKAGSVVIAGD 60
DB 1 MKFVLVAVLAYLIGCINNAIYFTKTRIDIRNYGSGNAGATNVLRLGPKAAAPVFLD 60
QY 61 ILKGTALPFLMHVDIHPLLAGVFAVLGHVFFIFAKFGKGAATSGGVLLFYAPLFF 120
DB 62 LKGTALPFLMHVDIHPLLAGVFAVLGHVFFIFAKFGKGAATSGGVLLFYAPLFF 120
DB 67 LKGVLAIALVYVLTFFASNQNLIPTVNIELWQPLVTLAGIAAILGHSKSIPLGFTGGK 126
QY 104 AVATSGGVLLFYAPLFFITMVAFFIFLYLTKFVSLSSMLTGITVIYFFVHD--TYLL 161
DB 127 SVATSLGLTAMNQVGLATFCGVFAVVVAISRVLSLSIMGAIAVSIWVLPLOQPLPIL 186
QY 162 IVVTLTITFIVYHRANKRIINTEPKV 190
DB 187 FGIA-GGLYVILRHRSNIERLLAGTEPKI 214

RESULT 10
QY 10 Q8UFU1 PRELIMINARY; PRT; 205 AA.
AC Q8UFU1
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Hypothetical protein Atul306.
GN ATU1306 OR AGR_C_2402.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
```


Db 5 LFPILGYFIGSIPFSLPKLWLGIDVRKVGSGNKGATNAIRTTGPAVGICLLLDALKG 64
OY 65 TLATALPFLMHVDIHL-LAGVPAVLGHVPPFAKFKGKAVATSGVLLFYAPLLFITVM 123
Db 65 FFPVFTITTSKIVSTATAITLGHDFPIPMFKGKGKAVASTIGLIFCLSWPTGLVF 124
OY 124 VAVFFILYLTFRVSLSSMLTGIYTVISFFV--HDTYLLIVVTLTIFVIYHRANIKR 181
Db 125 TLTMLVIMLTAKYASGLSVALYVSALLGKGYDTGMLFLI--LAVLSTLHSENIOR 182
OY 182 IINKTEPKV 190
Db 183 LLNGTERKV 191
RESULT 13
O9ZAF3 PRELIMINARY; PRT; 199 AA.
ID AC O9ZAF3
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 21.3 kDa protein.
OS Thermus thermophilus.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB27;
RX MEDLINE=20461243; PubMed=11004195;
RA Sanchez R., Roovers M., Glansdorff N.;
RT "Organization and expression of a thermus thermophilus arginine
RT cluster: presence of unidentified open reading frames and absence of a
RT Shine-dalgarno sequence.";
RL J. Bacteriol. 182:5911-5915(2000).
DR EMBL; Y18353; CAA77137.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein.
SQ SEQUENCE 199 AA; 21331 MW; A3BFB1652F8F206B CRC64;
Query Match 31.4%; Score 301.5; DB 2; Length 199;
Best Local Similarity 38.0%; Pred. No. 6.8e-18;
Matches 71; Conservative 36; Mismatches 75; Indels 5; Gaps 4;
OY 7 IILAYLIGSIPGLIVGKLGIDIREHSGNLGATNAFTLGVKAGSVVIAGDILKGTPL 66
Db 8 LLLAYLFGSIPAGVLVAR--TYGVDIRKVGSGNIGATNVLRLALGWPALVVAFFDFVKGI 66
OY 67 ATALPFLMHVDIHL-LAGV--FAVLGHVPPFAKFKGKAVATSGVLLFYAPLLFITVM 124
Db 67 AVLVARAFGLS--DMLGGVLMVLAHLYSVFLRFRGKGKAVATSGFTLLFLDPAALATWF 125
OY 125 AVFFILYLTFRVSLSSMLTGIYTVISFFV--HDTYLLIVVTLTIFVIYHRANIKRII 183
Db 126 PIGLSVILLTRYVSGTGGVAAVFLSLALGRPLWEVATVFLMALLIFWTHRENKLRLQ 185
OY 184 NKTEPKV 190
Db 186 ECTERRL 192
RESULT 14
O9ZQL7 PRELIMINARY; PRT; 203 AA.
ID AC O9ZQL7
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical transmembrane protein SMC01362.
CN R01302 OR SMC01362.

OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol W., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591786; CAC45881.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 203 AA; 21133 MW; 69BD0C43A30017D9 CRC64;
Query Match 31.4%; Score 301.5; DB 16; Length 203;
Best Local Similarity 35.9%; Pred. No. 6.9e-18;
Matches 66; Conservative 42; Mismatches 75; Indels 1; Gaps 1;
OY 7 IILAYLIGSIPGLIVGKLGIDIREHSGNLGATNAFTLGVKAGSVVIAGDILKGTPL 66
Db 17 LVFGYLLGSIPGLILTRMAGLDVVRKIGSGNIGATNVLTRNKLAAATLFLDALKGTA 76
OY 67 ATALPFLMHVDIHL-LAGVFAVLGHVPPFAKFKGKAVATSGVLLFYAPLLFITVM 126
Db 77 AAAYASYGVGEA-GIAAGFAAFLGHLFFPWLVSFRGKGKAVATGIVLLGLMPVVLFFRAI 135
OY 127 FFIFLYLTKFVSLSSMLTGIYTVISFFVHDTYLLIVVTLTIFVIYHRANIKRIINT 186
Db 136 WLAMAKITRYSLSALVATAVPIALYAAGNGKVGALFAVMTAIAWIKHRANIORLLSGT 195
OY 187 EPKV 190
Db 196 ESRI 199
RESULT 15
O8YC64 PRELIMINARY; PRT; 201 AA.
ID AC O8YC64
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
CN BMEI10668.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorte R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009702; AAL53910.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.

Thu Dec 26 09:18:39 2002

```

KW Complete proteome.
SQ SEQUENCE 201 AA; 20507 MW; 105CA44587BB4CA1 CRC64;

Query Match 31.4%; Score 301; DB 16; Length 201;
Best Local Similarity 39.1%; Pred. No. 7.5e-18;
Matches 75; Conservative 33; Mismatches 78; Indels 6; Caps 3;

Qy 1 MLIALLIILAYLGSIPIGKLVCKLAKGIDIREHSGNLGATNAFRTLGVKAGSVVIAGD 60
Db 10 MLIGAL-IFGVVLGSIPIGKLVCKLAKGIDIREHSGNLGATNAFRTLGVKAGSVVIAGD 68
Qy 61 ILKGTLAT--ALPFLMHVDIHPHLLAGVFAVLGHVFFIFARFKGKGVATSGGVLLFYAPL 118
Db 69 ALKGTAAALIAAHFGQNAI---AAGFCATFGLFPVWIGFKGKGVATLVGLIGLAWA 125
Qy 119 LFIWVAVFFIFLYLTKFVLSLMTGIYTVIYSEFFVHDYLLIVVTLITFVIYHRAN 178
Db 126 CALVFAAAWIVTALLARYSSLSALVASLVVPIALYSRGNQALAAALFAIMTVIVFIKHRAN 185
Qy 179 IKRIINKTEPKV 190
Db 186 ISRLINGTESKI 197

```

Search completed: December 26, 2002, 01:16:20
Job time : 624.209 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 00:49:20 ; Search time 121.219 Seconds
(without alignments)
46.846 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALLLIAYLIGSIPSGL.....RHRANKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	193	4	US-09-222-938A-70
2	476.5	49.6	204	4	US-09-134-001C-5239
3	275.5	28.7	114	4	US-09-222-938A-13
4	93	9.7	489	4	US-09-134-001C-5241
5	90	9.4	513	4	US-09-097-889-15
6	84.5	8.8	203	4	US-09-134-001C-5565
7	83.5	8.7	496	4	US-09-134-001C-3001
8	82	8.5	433	4	US-09-134-001C-3085
9	81.5	8.5	320	1	US-08-465-980-2
10	81.5	8.5	320	2	US-09-053-303-2
11	81.5	8.5	320	4	US-09-605-785-527
12	81.5	8.5	320	4	US-09-439-313-527
13	81.5	8.5	320	4	US-09-339-115-2
14	81.5	8.5	320	5	PCT-US95-07093-2
15	81	8.4	264	4	US-09-134-001C-3780
16	81	8.4	413	4	US-09-134-001C-3702
17	80.5	8.4	279	4	US-09-134-001C-4900
18	80	8.3	372	4	US-09-134-001C-3887
19	80	8.3	396	4	US-09-134-001C-4443
20	79.5	8.3	283	4	US-09-134-001C-5491
21	79.5	8.3	371	4	US-09-134-001C-4502
22	79	8.2	412	4	US-09-134-001C-3949
23	79	8.2	538	4	US-09-134-001C-4633
24	78.5	8.2	283	4	US-09-134-001C-3623
25	78	8.1	348	1	US-08-176-126B-4
26	78	8.1	348	2	US-08-669-435-4
27	78	8.1	348	5	PCT-US94-14431A-4

28 77.5 8.1 273 4 US-08-936-165A-395
29 77.5 8.1 1212 4 US-09-268-866-2
30 77.5 8.0 405 4 US-09-134-001C-3496
31 76.5 8.0 243 4 US-09-134-001C-2961
32 76.5 8.0 478 4 US-09-134-001C-5065
33 76.5 8.0 482 4 US-09-134-001C-4309
34 76.5 8.0 502 4 US-09-134-001C-5674
35 76 7.9 87 4 US-09-134-001C-4986
36 76 7.9 761 4 US-09-235-451-4
37 76 7.9 800 4 US-09-134-001C-5655
38 75.5 7.9 472 4 US-09-354-129-2
39 75.5 7.9 472 4 US-09-504-357-2
40 75.5 7.9 554 4 US-09-134-001C-5109
41 75 7.8 242 4 US-08-980-832-32
42 74.5 7.8 442 4 US-09-134-001C-3444
43 74.5 7.8 462 2 US-08-898-976-2
44 74.5 7.8 462 2 US-08-898-976-4
45 74.5 7.8 477 4 US-09-134-001C-3487

ALIGNMENTS

RESULT 1
US-09-222-938A-70
; Sequence 70, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222.938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 70
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-70

Query Match 100.0%; Score 960; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.6e-96;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLIALLLIAYLIGSIPSGLIVGKLAGKIDIREHSGNLGATNAPRTIGVRAGSVVIAGD 60
Db 1 MLIALLLIAYLIGSIPSGLIVGKLAGKIDIREHSGNLGATNAPRTIGVRAGSVVIAGD 60
Qy 61 ILKGTALATAPFLMHVDIHPHLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLLF 120
Db 61 ILKGTALATAPFLMHVDIHPHLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLLF 120
Qy 121 ITWAVFFIFLYLKFKVSLSSMLTGIVTVIYVSFFVHDYTLIVVTLLTIFVIYHRANIK 180
Db 121 ITWAVFFIFLYLKFKVSLSSMLTGIVTVIYVSFFVHDYTLIVVTLLTIFVIYHRANIK 180
Qy 181 RIINKTEPKVKWL 193
Db 181 RIINKTEPKVKWL 193

RESULT 2
US-09-134-001C-5239
; Sequence 5239, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS


```

; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-097-889-15

```

Query Match 9.48; Score 90; DB 4; Length 513;
Best Local Similarity 24.0%; Pred. No. 0.13;
Matches 50; Conservative 34; Mismatches 80; Gaps 10;
Indels 44;

Qy	60	10	AYLIGSPISGLIYCKLAKGIDIDREHSG-----NLGATNAFTLGVKAGSVVIAGD	60
Db	308	ATMIIAIPITGVKVSFWLATE-----HGSNMKWSAAVILWALGFIELF-TVGGLTG-IVLANS	361	
Qy	61	ILKGTLTATLPEFLMHVDIHPLLAGVFAVLG---HVPEPIFAKEKGGKAVATSGGVLLFYAP	117	
Db	362	SLDIVLHDTYYVVAHFHYVLSMGAVFALMGGFIHWPFLEFSGYTLDOT-----YAK	411	
Qy	118	LLFITMVAVFPIFYLT----KEVLSSTM-----LTGIYTVIYSPFVHDTYLLIVVTL	167	
Db	412	IHF---TFIFGNVLTFFPQHFLGLSGMPRRYSIDPDATYTNWLLSSVGSFISLTAVML	467	
Qy	168	TIFVIYHRHRAIKRIINKTEP--KVWKL	193	
Db	468	MIFMIWEAFASKRKVLVMEEPSNNLEWL	495	

RESULT 6
US-09-134-001C-5565
; Sequence 5565, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

```

, FILE REFERENCE: GTC-0007
, CURRENT APPLICATION NUMBER: US/09/134,001C
, CURRENT FILING DATE: 1998-08-13
, PRIOR APPLICATION NUMBER: US 60/064,964
, PRIOR FILING DATE: 1997-11-08
, PRIOR APPLICATION NUMBER: US 60/055,779
, PRIOR FILING DATE: 1997-08-14
, NUMBER OF SEQ. ID NOS: 5674
, SEQ ID NO 5565
, LENGTH: 203
,

```

; LIFE: PRI
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-5565

Query Match	8.8%	Score	84.5	DB	4	Length	203
Best Local Similarity	23.0%	Pred. NO.	0.15				
Matches	41	Conservative	23	Mismatches	31	Indels	83
						Gaps	10

Oy	85	VFAVLGHVF--PIPAKPKGKGKA-----VATSGG 110
Db	2	IYRVLIEHGHLVNKLAKGKADLVITSPRKVT SQQYAITQSGLQGFIITIGG 61
Oy	111	VLLPYAPLFTT-MVAV-----PFIFYL-----LTKFVSLSMUTGI 146
		: :
Db	62	YLM--PPLMFELTLGVSIHYQPSFITYIFLYFIYFYFFITSRLKSLPIVAILSSLLYL 119
Oy	147	YTVIYSFVVDH----TYILLIVTLL-----TIFVIYHRANIKRIINKTEPKVKW 192
		: : : :
Db	120	KPDQHWFIDYIVLSVSHFILGVLLGEILASSWTF-----RLTFOR-----PKPSW 166

RESULT 7
US-09-134-001C-3001
; Sequence 3001, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS AUREUS INFECTION.
TITLE OF INVENTION: EPIDERMIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007

```

CURRENT APPLICATION NUMBER: US 09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3001
LENGTH: 496

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3001

```

Query Match	8.7%	Score	83.5	DB	4	Length	496
Best Local Similarity	24.7%	Pred. No.	0.64				
Matches	48	Conservative	33	Mismatches	68	Indels	45
						Gaps	10

[illegible]

QY 160 LLiVVTLTIFVIY 173
: | : : | : : |
DB 421 IGEIASIYTIWIVY 434

RESULT 8

US-09-134-001C-3085
; Sequence 3085, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION

```

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

```

CURRENT APPLICATION NUMBER: US/09/134.0010
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3089
; LENGTH: 433

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3085

```

Query Match: 8.5%: Score 82: DB 4: Length 433:

BEST LOCAL SIMILARITY 21.8%; FREQ. NO.: 0.76;
Matches 57; Conservative 35; Mismatches 68;
Indels 102; Gaps 14

Qy	2	LIALLIILAYL-----IGSIPGSL-----IVGKLAKG-IDIREHGSN LGATNAF- 45
		: : : : : : : : :
Dø	109	VYFELIVL-YLCLNPKNMIDRIGSLTPLLLITIVAMIKGFDVDSGH-SSNYGMTWYH 166

Qy 46 -----RTLGVKAGSVIAGDILKGTLAT-ALP 71
: |: : : |: : : ||
Db 167 SNLSGFSOGFTOGVLTMDATASIAFNSHVVNAIKTTGIGHADKIFKQTIAGLIAAATLV 238

QY 72 FLM-----HV-----DIHPLLAGVFAVLGHVFPPIPAKFKGKAVA-- 106
 Db 227 FIVISLGYIGNINIPSDTLKELKAKDQNGITLTTMTATKG--FGTFCKYLLGIIVLSLA 284
 QY 107 ----TSGVLL----FYAPLFTTMVAVFFIPLYLTKFV-----SLSSMLTGIVTVIYS 152
 Db 285 CLTTAGGLIVSVSEFHRILPKYKVFVIFILVSFILANQGLNSVIKMSVPVLSVIYP 344
 QY 153 FFVHDYLLIVVTLTIFVIYR 174
 Db 345 -----VAITVILLILIAIR 357

RESULT 9

US-08-465-980-2
 ; Sequence 2, Application US/08465980
 ; Patent No. 5756309
 ; GENERAL INFORMATION:
 ; APPLICANT: Soppet, Daniel R.
 ; APPLICANT: Li, Yi
 ; APPLICANT: Rosen, Craig A.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
 ; ADDRESSEE: STUART & OLSTEIN
 ; STREET: 6 Becker Farm Road
 ; CITY: Roseland
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07068

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/465.980

FILING DATE: 06-JUN-1995
 CLASSIFICATION: 536
 NAME: Ferraro, Gregory D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 325800-446

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:

LENGTH: 320 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-465-980-2

Query Match 8.5% Score 81.5; DB 1; Length 320;
 Best Local Similarity 23.8%; Pred. No. 0.59;
 Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNMGATNAPFTLGKAGSVVIAGDILKG 64
 Db 145 IVAVVRSGLFFFPFLPLIKRL-----AFCHSNVLSHSCVHODVMKL 186
 QY 65 TLATALP-----FLMHVDIHLPLAGVFAVLGHVFPPIPAKFKGKAVATS-----GG 110
 Db 187 AVADTLPNVYGLTALLVMGVDVMFISLSYFLIIRTVLQPLPSKSERAKAFGTCVSHIGV 246
 QY 111 VLLFVAPLFTTMVAVFFIPLYLTKFVLSLSSMLTGIVTVIYSFFVHDYLLI 162
 Db 247 VLAFTVPLIGLSVWHRF-----GNSLHPIVRVV-----MGDIYLLL 282

RESULT 10

US-09-053-303-2
 ; Sequence 2, Application US/09053303
 ; Patent No. 5948890
 ; GENERAL INFORMATION:
 ; APPLICANT: Soppet, Daniel R.
 ; APPLICANT: Li, Yi
 ; APPLICANT: Rosen, Craig A.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
 ; ADDRESSEE: STUART & OLSTEIN
 ; STREET: 6 Becker Farm Road
 ; CITY: Roseland
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07068

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/053.303

FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/465.980

FILING DATE: 06-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferraro, Gregory D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 325800-446

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:

LENGTH: 320 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-053-303-2

Query Match 8.5% Score 81.5; DB 2; Length 320;
 Best Local Similarity 23.8%; Pred. No. 0.59;
 Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNMGATNAPFTLGKAGSVVIAGDILKG 64
 Db 145 IVAVVRSGLFFFPFLPLIKRL-----AFCHSNVLSHSCVHODVMKL 186
 QY 65 TLATALP-----FLMHVDIHLPLAGVFAVLGHVFPPIPAKFKGKAVATS-----GG 110
 Db 187 AVADTLPNVYGLTALLVMGVDVMFISLSYFLIIRTVLQPLPSKSERAKAFGTCVSHIGV 246
 QY 111 VLLFVAPLFTTMVAVFFIPLYLTKFVLSLSSMLTGIVTVIYSFFVHDYLLI 162
 Db 247 VLAFTVPLIGLSVWHRF-----GNSLHPIVRVV-----MGDIYLLL 282

RESULT 11

US-09-605-785-527
 ; Sequence 527, Application US/09605785
 ; Patent No. 6321716
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqi

```

; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 527
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-527

Query Match      8.5%; Score 81.5; DB 4; Length 320;
Best Local Similarity 23.8%; Pred. No. 0.59;
Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRSGLFFPLPLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFFIFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQPLSKSERAKAFGTCVSHIGV 246

QY 111 VLLFYAPLLFITMVAFFIFLYLTKFVLSLSMLTGIYTVIYFFVHDYLLI 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 VLAFYVPLIGLSVVRHF-----GNSLHPVIRVW-----MGDIYLLL 282

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 527
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-313-527

Query Match      8.5%; Score 81.5; DB 4; Length 320;
Best Local Similarity 23.8%; Pred. No. 0.59;
Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRSGLFFPLPLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFFIFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQPLSKSERAKAFGTCVSHIGV 246

QY 111 VLLFYAPLLFITMVAFFIFLYLTKFVLSLSMLTGIYTVIYFFVHDYLLI 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 VLAFYVPLIGLSVVRHF-----GNSLHPVIRVW-----MGDIYLLL 282

```

```

Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRSGLFFPLPLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFFIFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQPLSKSERAKAFGTCVSHIGV 246

QY 111 VLLFYAPLLFITMVAFFIFLYLTKFVLSLSMLTGIYTVIYFFVHDYLLI 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 VLAFYVPLIGLSVVRHF-----GNSLHPVIRVW-----MGDIYLLL 282

RESULT 13
US-09-339-115-2
; Sequence 2, Application US/09339115
; Patent No. 6372891
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Li, Yi
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/339,115
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 09/053,303
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-446
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-339-115-2

Query Match      8.5%; Score 81.5; DB 4; Length 320;
Best Local Similarity 23.8%; Pred. No. 0.59;
Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRSGLFFPLPLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFFIFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQPLSKSERAKAFGTCVSHIGV 246

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 01:16:25 ; Search time 80.8128 Seconds
(without alignments)
41.350 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALLLIAYLIGSIPSGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 109717 seqs, 17314136 residues

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	193	12	US-10-068-080-3
2	468	48.8	213	10	US-09-815-242-13329
3	468	48.8	213	10	US-09-815-242-13336
4	468	48.8	213	12	US-10-068-080-1
5	465.5	48.5	202	10	US-09-815-242-5305
6	465.5	48.5	202	10	US-09-815-242-12137
7	464.5	48.4	202	10	US-09-823-246-2
8	261.5	27.2	205	10	US-09-815-242-11713
9	257.5	26.8	203	10	US-09-815-242-13757
10	255	26.6	199	10	US-09-815-242-10997
11	252	26.2	189	10	US-09-815-242-11799
12	250.5	26.1	205	10	US-09-815-242-10303
13	230	24.0	220	10	US-09-815-242-11614
14	228	23.8	262	10	US-09-815-242-11453
15	93.5	9.7	299	10	US-09-393-634-35
16	92.5	9.6	451	10	US-09-815-242-10547
17	90	9.4	513	10	US-09-098-079-15
18	88	9.2	467	10	US-09-767-041-23
19	85	8.9	933	10	US-09-815-242-11817

20	82.5	8.6	822	10	US-09-824-734-3	Sequence 3, Appli
21	81.5	8.5	320	9	US-09-968-033C-4	Sequence 4, Appli
22	81.5	8.5	320	9	US-10-012-896-527	Sequence 527, App
23	81.5	8.5	320	9	US-09-895-793-527	Sequence 527, App
24	81.5	8.5	320	9	US-09-895-814-527	Sequence 527, App
25	81.5	8.5	320	10	US-09-759-143-527	Sequence 527, App
26	81.5	8.5	320	10	US-09-730-018-7	Sequence 7, Appli
27	81.5	8.5	320	10	US-09-780-669-527	Sequence 527, App
28	81.5	8.5	320	10	US-09-822-827-527	Sequence 527, App
29	81.5	8.5	320	10	US-09-886-055-83	Sequence 83, Appli
30	81.5	8.5	320	12	US-10-079-719-2	Sequence 2, Appli
31	81.5	8.5	320	9	US-09-968-033C-2	Sequence 2, Appli
32	79.5	8.3	320	9	US-09-815-242-5660	Sequence 5660, Ap
33	79	8.2	439	10	US-09-815-242-12272	Sequence 12272, A
34	79	8.2	484	10	US-09-815-242-13341	Sequence 13341, A
35	79	8.2	650	10	US-09-815-242-5678	Sequence 5678, Ap
36	78.5	8.2	430	10	US-09-860-232A-8	Sequence 8, Appli
37	78.5	8.2	468	10	US-09-939-980-395	Sequence 395, App
38	77.5	8.1	273	10	US-09-815-242-11566	Sequence 11566, A
39	77.5	8.1	391	10	US-09-815-242-10592	Sequence 10592, A
40	77.5	8.1	417	10	US-09-895-913A-146	Sequence 146, App
41	77.5	8.1	435	9	US-10-117-417-4	Sequence 4, Appli
42	77.5	8.1	445	9	US-10-117-417-15	Sequence 15, Appli
43	77.5	8.1	513	10	US-09-741-148A-2	Sequence 2, Appli
44	77.5	8.1	533	10	US-09-815-242-11612	Sequence 11612, A
45	77.5	8.1	533	10	US-09-815-242-11612	Sequence 11612, A

ALIGNMENTS

RESULT 1
US-10-068-080-3
; Sequence 3, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritzt, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-3

Qy	1	MLIALLLIAYLIGSIPSGLVKLGKIDIREHSGNLGNATNFRILGVKAGSVIAGD	60
Db	1	MLIALLLIAYLIGSIPSGLVKLGKIDIREHSGNLGNATNFRILGVKAGSVIAGD	60
Qy	61	ILKCTLTALPFLMHVDIHPHLLAGVAVLGHVFFIFAKFGKAVATSGGVLLFVAPLLF	120
Db	61	ILKCTLTALPFLMHVDIHPHLLAGVAVLGHVFFIFAKFGKAVATSGGVLLFVAPLLF	120
Qy	121	ITWVAVFFIPLYLTKFVSLSSMLTGIYTVIYFVFDHTYLLIVVTLTIFVIYRHRANIK	180
Db	121	ITWVAVFFIPLYLTKFVSLSSMLTGIYTVIYFVFDHTYLLIVVTLTIFVIYRHRANIK	180
Qy	181	RIINKTEPKVKWL	193
Db	181	RIINKTEPKVKWL	193

Query Match 100.0%; Score 960; DB 12; Length 193;
Best Local Similarity 100.0%; Pred. No. 2.6e-83;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

US-09-815-242-13329
; Sequence 13329, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13329

; LENGTH: 213

; TYPE: PRF

; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13329

Query Match 48.8%; Score 468; DB 10; Length 213;
Best Local Similarity 48.0%; Pred. No. 4.4e-37;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

QY 1 MLIALIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNAPRTGLGVAGSVIAGD 60

DB 1 MITIVLLIAYLIGSIPSLVGLKAGIDIREHSGNLGATNAPRTGLGVAGSVIAGD 60

QY 61 ILKGTALATLPFLMHVD-IHPLLAGVFAVLGHVFPFAKFGKAVATSGVLLFYAPLL 119

DB 61 FFKGTATLPIIFHLQGVSLPIFGLLAVIGHTTPIFAGKGVATSGVIFGFAPIF 120

QY 120 FITWAVFFIYLTKFVSLSSMLTGIVTVI-----YSFFV--HDTYLLIVVTLTIF 170

DB 121 CLYLAIIFGALYGLSMISLSSVTASIAAIGVLLFPLFGFILSNYDSLFTAILALASL 180

QY 171 VIYHRANKRIINKTEPKVKW 192

DB 181 IIRHKDNIARIKKNKTENLVPM 202

RESULT 3

US-09-815-242-13536
; Sequence 13536, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13536

; LENGTH: 213

; TYPE: PRF

; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13536

Query Match 48.8%; Score 468; DB 10; Length 213;

Best Local Similarity 48.0%; Pred. No. 4.4e-37;

Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

QY 1 MLIALIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNAPRTGLGVAGSVIAGD 60

DB 1 MITIVLLIAYLIGSIPSLVGLKAGIDIREHSGNLGATNAPRTGLGVAGSVIAGD 60

QY 61 ILKGTALATLPFLMHVD-IHPLLAGVFAVLGHVFPFAKFGKAVATSGVLLFYAPLL 119

DB 61 FFKGTATLPIIFHLQGVSLPIFGLLAVIGHTTPIFAGKGVATSGVIFGFAPIF 120

QY 120 FITWAVFFIYLTKFVSLSSMLTGIVTVI-----YSFFV--HDTYLLIVVTLTIF 170

DB 121 CLYLAIIFGALYGLSMISLSSVTASIAAIGVLLFPLFGFILSNYDSLFTAILALASL 180

QY 171 VIYHRANKRIINKTEPKVKW 192

DB 181 IIRHKDNIARIKKNKTENLVPM 202

RESULT 4

US-10-068-080-1
; Sequence 1, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 213
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-1

Query Match 48.8%; Score 468; DB 12; Length 213;
Best Local Similarity 48.0%; Pred. No. 4.4e-37;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

QY 1 MLIALIILAYLIGSIPSGLVGKLAGIDIREHSGNLGATNAPRTLGKAGSVVIAGD 60
 Db 1 MITIVLLIAYLIGSIPSGLVGKLAGIDIREHSGNLGATNAPRTLGKAGSVVIAGD 60
 QY 61 ILKGTALATLPFLM--HVD-----IHPLLAGVFAVLGHVFPFAKFKGKAVATSGV 119
 Db 61 FPKGTLATLPFIIFHLQGVSPFLFGLLAVIGHTFPFIFAGFKGKAVATSGVIFGFAPIF 120
 QY 120 FITWAVFFIYLYKFKVSLSSMLTGIVTVI-----YSFV--HDYLLIVVTLITFV 170
 Db 121 CLYLAIIFFGALYGLSMISLSSVTASIAAVIGLVLLFPLFGFILTSDYSLIFAILALASL 180
 QY 171 VIYHRANKRIINTEPKVKW 192
 Db 181 IIRHRSNISRIFRGEEPKIKWM 202

RESULT 5
 ; Sequence 5305, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5305
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5305

Query Match 48.5%; Score 465.5; DB 10; Length 202;
 Best Local Similarity 46.5%; Pred. No. 7.1e-37;
 Matches 94; Conservative 40; Mismatches 59; Indels 9; Gaps 2;
 QY 1 MLIALIILAYLIGSIPSGLVGKLAGIDIREHSGNLGATNAPRTLGKAGSVVIAGD 60
 Db 1 MNIIVMLLSYLIAGFPSPGVLGKLFKKDIRQFGSGNTGATNSFRVLGRPAGFLVTFD 60
 QY 61 ILKGTALATLPFLM--HVD-----IHPLLAGVFAVLGHVFPFAKFKGKAVATSGV 111
 Db 61 IFKGTITVFFLWLPVHADGPISTFTNGLIVGLFAILGHVFPVYLFKFGGKAVATSGV 120
 QY 112 LLFYAPLLEITWAVFFIYLYKFKVSLSSMLTGIVTVIYFVFFVHDYLLIVVTLITFV 171
 Db 121 VLGVNPILLILAIIFFIYLYKFKVSLSSMLTGIVTVIYFVFFVHDYLLIVVTLITFV 180

QY 172 IYHRANKRIINTEPKVKWL 193
 Db 181 IIRHRSNISRIFRGEEPKIKWM 202
 RESULT 6
 ; Sequence 12137, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12137
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-12137

Query Match 48.5%; Score 465.5; DB 10; Length 202;
 Best Local Similarity 46.5%; Pred. No. 7.1e-37;
 Matches 94; Conservative 40; Mismatches 59; Indels 9; Gaps 2;
 QY 1 MLIALIILAYLIGSIPSGLVGKLAGIDIREHSGNLGATNAPRTLGKAGSVVIAGD 60
 Db 1 MNIIVMLLSYLIAGFPSPGVLGKLFKKDIRQFGSGNTGATNSFRVLGRPAGFLVTFD 60
 QY 61 ILKGTALATLPFLM--HVD-----IHPLLAGVFAVLGHVFPFAKFKGKAVATSGV 111
 Db 61 IFKGTITVFFLWLPVHADGPISTFTNGLIVGLFAILGHVFPVYLFKFGGKAVATSGV 120
 QY 112 LLFYAPLLEITWAVFFIYLYKFKVSLSSMLTGIVTVIYFVFFVHDYLLIVVTLITFV 171
 Db 121 VLGVNPILLILAIIFFIYLYKFKVSLSSMLTGIVTVIYFVFFVHDYLLIVVTLITFV 180
 QY 172 IYHRANKRIINTEPKVKWL 193
 Db 181 IIRHRSNISRIFRGEEPKIKWM 202
 RESULT 7
 ; Sequence 246-2
 ; Patent No. US2002058789A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Burnham, Martin K. R.

```
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Warren, Patrick V.
; APPLICANT: Sylvester, Daniel R.
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: ynes
; FILE REFERENCE: GM20001
; CURRENT APPLICATION NUMBER: US/09/823,246
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,496
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-823-246-2

Query Match      48.48; Score 464.5; DB 10; Length 202;
Best Local Similarity 46.58; Pred. No. 8.8e-37;
Matches 94; Conservative 40; Mismatches 59; Indels 9; Caps 2;

Qy 1 MLIALIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGVKAGSVIAGD 60
Db 1 MWIIVLLSLYLIGAFSGVIGLKFKKDIRQFSGNGTATNSFRVLGRPAGFLVTFD 60

Qy 61 ILKGTALATAPFLM--HVD-----IHPLLAGVFAVLGHVFFIPAKFKGKAVATSGV 111
Db 61 IFKGFITVFFPLVQLVHADGPISTFTFNTGLVGLFAILGHVYVYLVKFGGKAVATSGV 120

Qy 112 LLFYAPLFTIWMVAVFFFLYLTKFVSLSSMLTGTIYTVIYFFVHDYLLIIVTLLIFV 171
Db 121 VLVNPIILLIATIFFVLKIFRYSLASIVAAICCVIGSLIIOQYILLVVSFLVSIIL 180

Qy 172 IYHRANKRIINKTEPKVKW 193
Db 181 IIRHSNIARIFRGEPEKIKW 202

RESULT 8
US-09-815-242-11713
; Sequence 11713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13757
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13757

Query Match      26.88; Score 257.5; DB 10; Length 203;
Best Local Similarity 35.58; Pred. No. 2.4e-17;
Matches 70; Conservative 37; Mismatches 69; Indels 21; Caps 7;

Qy 6 LIIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGVKAGSVIAGDILKGT 65
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11713
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11713

Query Match      27.28; Score 261.5; DB 10; Length 205;
Best Local Similarity 35.08; Pred. No. 1e-17;
Matches 69; Conservative 39; Mismatches 68; Indels 21; Caps 7;

Qy 6 LIIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGVKAGSVIAGDILKGT 65
Db 8 LVLLAYLCGSISSAILVCRLAGLPDRDSGSGNGATNVLIGCGKGAVALIFDVLKGM 67

Qy 66 L-----ATALL---PFLMHVDIHLPLAGVFAVLGHVFFIPAKFKGKAVATSGVLLFYAPL 118
Db 68 LPVNGAWALGLTFPWLG-----LVAIAACVGHVHPVFFHFRGCKGVATAGAI---API 118

Qy 119 ---LFTIWMVAVFFFLYLTKFVSLSSMLTGTIYTVIYFFVHDYLLIIVTLLIFVIYRH 175
Db 119 GLDVTGYMAGTLLTLLISYSSLGATVSAIAPFYVMFKPOY-TFPVSNLSCLILLRH 177

Qy 176 RANIKRIINKTEPKVKW 192
Db 178 HDNIQRLWRROESKI-W 193

RESULT 9
US-09-815-242-13757
; Sequence 13757, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13757
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13757

Query Match      26.88; Score 257.5; DB 10; Length 203;
Best Local Similarity 35.58; Pred. No. 2.4e-17;
Matches 70; Conservative 37; Mismatches 69; Indels 21; Caps 7;

Qy 6 LIIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGVKAGSVIAGDILKGT 65
```



```
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10303
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10303
```

```
Query Match 26.1%; Score 250.5; DB 10; Length 205;
Best Local Similarity 33.2%; Pred. No. 1.1e-16;
Matches 64; Conservative 41; Mismatches 75; Indels 13; Gaps 6;

QY 6 LLIILAYLIGSPGLIVKLGAKGIDIREHSGNLGATNAPRTLGVKAGSVVIADILKGT 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 8 MLIIAYLCGSISSAILVCRGLCPDPTSGSGNPGATNVLRIIGKGAAGAAVAVLFDVLKGM 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 LATALPFLMHVDIHPLLAGFAV---LGHVFIPIAKFKGKAVATSGGVLLFYAPL---L 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 LPVWCAY--ELGVSPFWLGLIAIAACLGHIWPFVFGFGKGVATAFCAI---APIGHD 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 FITWAVAFIFLYLTKFVLSLSMLTGIVTVIYSFVHDVTLVLLVTLTIFVIYHRANI 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 TGVMAGTWLLTVLLSGYSLGAIVSALIAPIFYVMFKPQF-TFPVSMLSCLILLRHDNI 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 KRIINKTEPKVKW 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 ORLWRRQETKI-W 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 13
US-09-815-242-11614
; Sequence 11614, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11614
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11614
```

```
Query Match 24.0%; Score 230; DB 10; Length 220;
Best Local Similarity 31.6%; Pred. No. 1e-14;
Matches 65; Conservative 38; Mismatches 83; Indels 20; Gaps 6;

QY 5 LLIILAYLIGSPGLIVKLGAKGIDIREHSGNLGATNAPRTLGVKAGS-----VVI 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 14 IFTLLGLYLGIGIPGYALMKIFYGMDITKIGSGGIGATNVLRALQSKGVSNAKOMALLVL 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 AGDILKGTALATLPFLMHVDIH-PLLACVFAVLGHVPEPIFAKFKGKAVATSGGVLLFYA 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 74 ILELFKGMFAVFLSKLFLGDISLQMVAVIASILGHCHYSPFLNFGKGVSTINGSVVLLI 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 PLAFITWAVAFIFLYLTKFVLSLSML-TGIYTVIYSF--FVH--DTYLLI-----VV 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 134 PIESLIGLTVFWFGVKLVKISSLASILGVTATVLIFFVPMHIPDSVNLKKEVGTQTPM 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 165 TLLTIFVIYHRANIKRIINKTEPKV 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 194 VLIFITLKHAGNIFNLLTGKCKV 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 14
US-09-815-242-11453
; Sequence 11453, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11453
; LENGTH: 262
```

; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11453

Query Match 23.8%; Score 228; DB 10; Length 262;
Best Local Similarity 31.6%; Pred. No. 1.9e-14;
Matches 65; Conservative 38; Mismatches 83; Indels 20; Gaps 6;

QY 5 LLIILAYLIGSIPSLIVGLKAGIDIREHSGNLGATNAFTLGVKAGS-----VVI 57
DB : : | | | | | : : | : | : | | : | | : | :
56 IFTLLGLYLGIGIPGYALMKIFYGMDITKISGGIGATNVLRLQSKGVSNAKOMALLVL 115
QY 58 AGDILKGTALPFLMHVDIH-PLLAGVFAVLGHVPIFAKFKGKAVATSGGVLLFYA 116
DB : : | | | | | : : : : | | : | | | | : | :
116 ILDLFKGMFAVFLSKLGLDYSLOWVVAIASILGHCHYSPFLNFGGKGVSTIMGSVVLLI 175
QY 117 PLLEITWAVFFFLYLTKFVLSLSSML-TGIYTVIYSF--FVH--DTYLLI-----VV 164
DB : : : | : | : | | : | : | : | : | : | : | :
176 PIESLIGLTWVFFVKVGLKISSLASILGCVGTATVLIFFVPMHIPDSVNILKEVGTQTPM 235
QY 165 TLLTIFVIYHRANIKRIINKTEPKV 190
DB : : | : | : | : | : | : | : | : | : | :
236 VLIFFTLIRKHAIGNIFNLLAGKEKV 261

RESULT 15

US-09-393-634-35
; Sequence 35, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393, 634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human GR01, SF01
US-09-393-634-35

Query Match 9.7%; Score 93.5; DB 10; Length 299;
Best Local Similarity 23.2%; Pred. No. 0.092;
Matches 49; Conservative 41; Mismatches 80; Indels 41; Gaps 10;

QY 5 LLIILAYLIGSIPSLIVGLKAGIDIREHSGS-----GNLGNATNAFTLGVKAGSVV 56
DB : : | : | : | : | : | : | : | : | : | :
11 LLAIVQLGIFTNGIIV--VVGIDLIKHKMKAPLDLLSCLAVSRIFLQLFTFYVNV 68
QY 57 I---AGDILKGTALPFLMHVDIHPLLA---GVF-----AVLGHVPIFAKFKGKAV 105
DB : : | : | : | : | : | : | : | : | : | :
69 VIFFIEFIMCSANCAILLFNELEL--WLATWLGVFYCAKVASVRHPLFLWLKMRISKLV 126
QY 106 ATSGVLLFYAPLFTIMWAV-----FFILYLTKFVLSLSSMLTGIYV---IYSFFV 155
DB : : | : | : | : | : | : | : | : | : | :
127 P-----WMLGSLLYVSMICVFSKYGACFWVPYFLRAFFFSQNTIQRKEDTLAQIFSVA 181
QY 156 HDTYLLIV---VTLLTIFVIYHRANIKRII 183
DB : : | : | : | : | : | : | : | : | : | :
182 EFSVPLLIFFLFAVLLIFSLGRHTQMRNTV 212

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:13:55 ; Search time 10196.6 Seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-4
Perfect score: 582
Sequence: 1 atgtaattgctttattgat.....ctaaagtaaaatggttataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_nam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	582	100.0	6595	1	AF024713	Bacillus
c 2	582	100.0	26170	1	BC170DEGR	273234 B.subtilis
c 3	582	100.0	233780	1	BSUB0010	299113 Bacillus su
c 4	176	30.2	195269	6	AX417035	AX417035 Sequence
c 5	176	30.2	349980	6	AX417044	AX417044 Sequence
c 6	171.8	29.5	250050	1	AL591978	AL591978 Listeria
c 7	168.6	29.0	7563	1	AF084044	AF084044 Listeria
c 8	121.4	20.9	9839	1	AB010565	AB010565 Fusobacte
c 9	117	20.1	25021	1	AB028896	AB028896 Streptoco
c 10	114	19.6	771	1	SG0236899	AJ236899 Streptoco
c 11	114	19.6	11545	1	AE005539	AE005539 Streptoco
c 12	114	19.6	11573	1	AE010024	AE010024 Streptoco
c 13	114	19.6	53354	1	AE014149	AE014149 Streptoco
c 14	108.2	18.6	3014	1	AF269800	AF269800 Staphyloc
c 15	108.2	18.6	3014	6	AX145118	AX145118 Sequence
c 16	108.2	18.6	4105	1	AF269437	AF269437 Staphyloc
c 17	108.2	18.6	4105	6	AX144757	AX144757 Sequence
c 18	106.6	18.3	594	1	AX144133	AX144133 Sequence
c 19	106.6	18.3	13211	1	AE006332	AE006332 Lactococc
c 20	101.6	17.5	10828	1	AE008451	AE008451 Streptoco
c 21	101.6	17.5	248254	2	SPNEU1903	AL449925 Streptoco
c 22	100	17.2	6171	6	BD003717	BD003717 Polynucle
c 23	100	17.2	6812	1	SPPARCETP	267739 Streptococc
c 24	100	17.2	10624	1	AE007390	AE007390 Streptoco
c 25	97.8	16.8	11711	1	AE013117	AE013117 Thermoana
c 26	89	15.3	295350	1	AP004826	AP004826 Staphyloc
c 27	89	15.3	303750	1	AP003133	AP003133 Staphyloc
c 28	89	15.3	346900	1	AP003362	AP003362 Staphyloc
c 29	72.8	12.5	14651	1	AE012788	AE012788 Chlorobiu
c 30	67.2	11.5	12639	1	AE009092	AE009092 Agrobacte
c 31	67.2	11.5	14140	1	AE008058	AE008058 Agrobacte
c 32	67.2	11.5	15663	1	AE001796	AE001796 Thermotog
c 33	64.6	11.1	13750	1	AE000701	AE000701 Aquifex a
c 34	62.8	10.8	146174	1	D90910	D90910 Synecocyst
c 35	61.8	10.6	8634	1	AF281816	AF281816 Acholepla
c 36	57.2	9.8	10530	1	AE013956	AE013956 Yersinia
c 37	55.2	9.5	1657	6	AX416329	AX416329 Sequence
c 38	54.2	9.3	342650	1	AP003582	AP003582 Nostoc sp
c 39	53	9.1	12216	1	AE006206	AE006206 Pasteurel
c 40	52.4	9.0	13785	1	AE005913	AE005913 Caulobact
c 41	52	8.9	293350	1	SME591786	AL591786 Sinorhizo
c 42	48.8	8.4	7218	6	I66494	I66494 Sequence 14
c 43	48	8.2	10177	1	U32713	U32713 Haemophilus
c 44	48	8.2	11198	1	AE004138	AE004138 Vibrio ch
c 45	47.4	8.1	16662	1	AE002138	AE002138 Ureaplasma

ALIGNMENTS

RESULT 1
AF024713/c
LOCUS AF024713 6595 bp DNA linear BCT 05-OCT-1999
DEFINITION Bacillus subtilis DNA topoisomerase IV subunits ParE (parE) and
ParC (parC) genes, complete cds.
ACCESSION AF024713
VERSION AF024713.1 GI:2558945
KEYWORDS Bacillus subtilis.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
REFERENCE 1 (bases 1 to 6595)
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
AUTHORS Huang W.M., Libbey J.L., van der Hoeven P. and Yu S.X.
TITLE Bipolar localization of Bacillus subtilis topoisomerase IV, an
enzyme required for chromosome segregation

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (8), 4652-4657 (1998)
 MEDLINE 98208596
 PUBMED 9539793
 REFERENCE 2 (bases 1 to 6595)
 AUTHORS Huang, W.M., Libbey, J.L., van de Hoeven, P. and Yu, S.X.
 TITLE Direct Submission
 JOURNAL Submitted (13-SEP-1997) Oncological Sciences, University of Utah, Salt Lake City, UT 84132, USA
 FEATURES Location/Qualifiers
 source 1..6595
 /organism="Bacillus subtilis"
 /strain="CB10, a derivative of 168"
 /db_xref="taxon:1423"
 /map="1800 kbp"
 2073..4040
 /gene="parE"
 /note="required for chromosome segregation"
 2073..4040
 /gene="parE"
 /note="a subunit of DNA topoisomerase IV"
 /codon_start=1
 /transl_table=11
 /product="ParE"
 /protein_id="AAB81614.1"
 /db_xref="GI:2558946"
 /translation="MARKQDFYEDNAIOVLEGLAEVNRKPRMVGISTDARGLHLVY
 EIVNSDVEIAGHGDHIIIVKHNDLSISVDGRGMPGKMGKLGKPTPEVILVLA
 GKGFGQGGYISGGLHGVASVNAALSEMLVYTERDGFVYQORFGRKPVTSLEKI
 GKTKGTGLTHFQDPMTFTTYNFTLSERLRESAFLLKGLKIELIDERNQREVF
 YVGEIAFAFVYLNEDKVLSEVSVFGEHRSIEVDFAFQNDGYSNLSFVNNVTQ
 KDGTGSEKATATRAFNEVARKVALLKNDLETDIDIRGSAISISVRIPEELQ
 FEGQTKLGTSEARSADVAIVSEQLAYFLENRDQATLLVKAIKASQAREARKAR
 EARSGRKKRSEATLSGLKTPAGSRPAKKNELVLEDGSAGGAKQORRRFQVPLD
 LRGVINTEKAKLADIKNKEINTIIHAIGGVGADFSIDIDYDKIIMTDVADLA
 HIQVILLTFFYRKPLFETGKGVFIAPLPYKVKSGSKKEIIIEYAWSDEMGDLVK
 VKGVTIORYKGLGEMNADQLWETMPESRTLVVRKIDDAARVRRVTTLMGDKVEP
 RKWKTEKNVAFGLDESNLENENLSVAEEV"
 4044..6464
 /gene="parC"
 /note="required for chromosome segregation"
 4044..6464
 /gene="parC"
 /note="a subunit of DNA topoisomerase IV"
 /codon_start=1
 /transl_table=11
 /product="ParC"
 /protein_id="AAB81615.1"
 /db_xref="GI:2558947"
 /translation="MSQPELPHDLPLLEEVIGDRFGYSKYIIQDRALPDARDGLKPVQ
 RRLVAMHTDNTGDKNFRKAATVGNVIGNYHPHGDSVYEAEMVRKSDHKNVRLI
 EMHNGNSIDGPPAARMYTERLSPHGLVEIDAVIKRIQMPSCVDELMDEVKGFDPFT
 MFPNLLVNGSTGISAGYATDIPPHLGEIDAVIKRIQMPSCVDELMDEVKGFDPFT
 GGIQGVGDIRKAYETGKGIIRKAEIETIRGREQIVITEIPFEVYNKANLVKMD
 EFRIDKVEGISEVERDRTGRVYIELKKEADAKGLNFKNTDLOITQYFNFWA
 IHNRRPMLKSPISLDVAVIGHQKEVTVNRSVLEQAKDRHHIIVEGLMKALSILDEVI
 ATIRSSDRKAKNNLIKAYEETEPQAEIVSLQYRLTNTDITALEAEELCKKIE
 ELESISNDKLLKVVITNSLAKAKKYADTRRSVEEIKELNLEWVASEDYVVT
 VTKDGLTKRSRSPFAASNGDQFGMKDTHOFEMNTDVLFLTNGKSYIYCPVH
 QLPDIRKWDGCHFSNLTIDRDTVKAIPKEFPDPSAYLLFTFKNGSVKKTETLYH
 KARYSKALVALNKGEDELIDVHTNGESQIFMATHLGYGLWFGEDVNVVGARAAG
 VKGINLKEDDFVSGEIIQQSDSVILFTORGAVKRMISLEFEKTSRAKRGVYMLRELK
 KNPVRVALFACGLBQRLMAETEGDKRELQTKELRTNDRYSNCSFFPDEESGKVT
 VNRHLHQE"
 BASE COUNT 2096 a 1261 c 1543 g 1695 t
 ORIGIN
 Query Match 100.0%; Score 582; DB 1; Length 6595;
 Best Local Similarity 100.0%; Pred. No. 2.le-137;
 Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTTAATTCCTTTATTTGGCTACTCTTATAGGAGCATTCATCTGGCTTA 60
 Db 1097 ATGTTAATTCCTTTATTTGGCTACTCTTATAGGAGCATTCATCTGGCTTA 1038

QY 61 ATTGTGGCAGCTTGCACAAAGGAATGATATTCGGGAGCAGCAAGCGCAACTTAGGC 120
 Db 1037 ATTGTGGCAGCTTGCACAAAGGAATGATATTCGGGAGCAGCAAGCGCAACTTAGGC 978
 QY 121 GCTACCAATGATTCCTGATACATTTGGGTGTAAGAGTGGTTCGTCATAGCCGAGAT 180
 Db 977 GCTACCAATGATTCCTGATACATTTGGGTGTAAGAGTGGTTCGTCATAGCCGAGAT 918
 QY 181 ATTTTGAAGGACATGGGCAACTGCATTCCTTTCTCATGCATGTTGATATTCACCCG 240
 Db 917 ATTTTGAAGGACATGGGCAACTGCATTCCTTTCTCATGCATGTTGATATTCACCCG 858
 QY 241 CTCTTCGAGGAGTCCTTTCGCGGTTTAGGCCACGCTTTCCCATCTTCGCAAAATTTAAA 300
 Db 857 CTCTTCGAGGAGTCCTTTCGCGGTTTAGGCCACGCTTTCCCATCTTCGCAAAATTTAAA 798
 QY 301 GGCGTAAAGCGGTGGCGACATCAGGAGCGTTTGTCTATTATTTAGCGACCCCTGTTATT 360
 Db 797 GGCGTAAAGCGGTGGCGACATCAGGAGCGTTTGTCTATTATTTAGCGACCCCTGTTATT 738
 QY 361 ATCAGATGTTGGGTATTCATCTTTTATTTACTTTGACTTAATTTGTTCTCTCTCA 420
 Db 737 ATCAGATGTTGGGTATTCATCTTTTATTTACTTTGACTTAATTTGTTCTCTCTCA 678
 QY 421 TCGATGTTAAAGGATCTATCTGTTATATATAGTTTCTTTCCTCATGATACGTATTTA 480
 Db 677 TCGATGTTAAAGGATCTATCTGTTATATATAGTTTCTTTCCTCATGATACGTATTTA 618
 QY 481 TTGATGTCGTTACCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 540
 Db 617 TTGATGTCGTTACCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 558
 QY 541 CGAATTATCAATAAAGACAGACCTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 582
 Db 557 CGAATTATCAATAAAGACAGACCTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 516
 RESULT 2
 LOCUS BC170DEGR 26170 bp DNA linear BCT 24-JUN-1998
 DEFINITION B.subtilis DNA (26.2 kb fragment; 170 degree region).
 ACCESSION Z73234
 VERSION Z73234.1 GI:1405443
 KEYWORDS aconitase; alst gene; ccdA gene; endo-1,4-beta glucanase;
 endo-1,4-beta xylanase; grIA gene; grIB gene; tlpA gene;
 transketolase; yneA gene; yneB gene; yneE gene; yneF gene; yneI
 gene; yneJ gene; yneK gene; yneL gene; yneP gene; yneQ gene; yneR
 gene; yneS gene; yneT gene; yneC gene; yneF gene; yneG gene; yneH gene.
 SOURCE Bacillus subtilis.
 ORGANISM Bacillus subtilis.
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (bases 1 to 26170)
 AUTHORS Rose, M. and Entian, K.D.
 TITLE New genes in the 170 degrees region of the Bacillus subtilis genome
 encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
 acid transporter
 JOURNAL Microbiology 142 (Pt 11), 3097-3101 (1996)
 MEDLINE 9712194
 PUBMED 8969507
 REFERENCE 2 (bases 1 to 26170)
 AUTHORS Rose, M.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUN-1996) Rose M., Johann Wolfgang
 Goethe-Universitaet Frankfurt, Institut fuer Mikrobiologie,
 Marie-Curie-Str. 9, Geb. N250 Frankfurt/M. GERMANY D-60439
 COMMENT Overlapping sequence: X87845 (bases 3016-6779).
 FEATURES Location/Qualifiers
 source 1..26170
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 82..399
 gene


```

239113.1  GI:2634090
.
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 233780)
Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
Bronss,R., Bourslet,L., Brans,A., Braun,M., Brignell,S.C.,
Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrarri,A.,
Faulder,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Gallizzi,A.,
Galleron,N., Ghim,S.Y., Glaser,M., Goffeau,A., Gollightly,E.J.,
Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Halech,J., Harwood,C.R.,
Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
Lardinot,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
Portetle,D., Porwollik,S., Prescott,A.M., Prescan,E., Pujic,P.,
Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadale,Y., Sato,T.,
Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzneger,T.,
Winters,P., Wiput,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
Yoshida,K., Yoshikawa,H.F., Zumbstein,E., Yoshikawa,H. and
Danchin,A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
9384377
2 (bases 1 to 233780)
Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES
SOURCE
Location/Qualifiers
1. 233780
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
75. .744
/gene="pkxA"
75. .692
/gene="pkxA"
/function="regulation of the polyketide synthase operon
(pkx)"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="CAB13591.1"
/db_xref="GI:2634091"
/db_xref="SPTREMBL:O34381"
/translation="MPKQIDHRRKRQIAETRWVILRGMGASARNIAKEAGLSLQ
ALRHYSTQDELLAFAMKLVQEKVTRDKIDAVRDLLPKKVLQILLEMVPTNEETI
EMEWFAFYARHKDKMFASHDGIIFSGMRNLTYLDESLLQKADKIDEARLYR"
LVDGLAHMLDPVRVKNKRIKRVIMQHVESICVEDTREQKRHP"
709. .744
/gene="pkxA"
882. .1559
terminator
gene

```

```

/genes="pkSB"
882..1559
/genes="pkSB"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13592.1"
/db_xref="GI:2634092"
/db_xref="SPTREMBL:O34769"
/translation="MNLTYKVHPKTRYOGWNYCYIIEDIVSRSAIVDPSPHLSKI
TTTSLSEAEKLAVALTHSHYDHNVDPLTKMFAOVNYSKKEIDYQFRCNRLISL
DDHOTISIGNTRAOCILHTAGTCYLFSEIFTGDTVFTGCGCIGEDDGSAAEM
FDSIQIRKISVSPHRYVPGHSFGKSGHSIKDLYOHNIYFQIDKKEYFYFKPRTRNQ
KGIFDFK"
1932..2798
/genes="pkSB"
1932..2798
/genes="pkSB"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13593.1"
/db_xref="GI:2634093"
/db_xref="SPTREMBL:O34825"
/translation="MITYVPGGSGOGKMGSLDFEFKELTDQADEILGYSIKRLCL
ENPYSNLNKTQTPALYVYNALSYLKKIRDEVKPDFVAGHSIGEYNALFAAEADF
ETGLQVLRKGLMSLLSNGGMAAIVGLNEEQVAKALKYHLHDVDIANVNPYQIVI
SGKDETEKAASLFEFTMETVYMLPLNVSGAFHSRYMNRKAKEFEFLHAFYSPPSI
PVISNYVAKPYTFEYFKKQTLADQINHSVKWTSISYLSLKKGHMEFEVPGNVLTGLI
HRIKKDAEMPR"
3302..4192
/genes="pkSD"
3302..4192
/genes="pkSD"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13594.1"
/db_xref="GI:2634094"
/db_xref="SPTREMBL:O34877"
/translation="MNEPLVFMFSGGQSYHMGKELFKENTVFRSMLEMDIAAARR
IGTSIVEEYHPGRKSDPDFDLSLFPALFMEYSLYKVLDEDRGYPDVYLGSVWEN
SCDKGMLALDLPOLLNDHPDLFNSLISINVDHNFVSGEDHIRKIMEDLKEQO
ILCOLLPVSVAHUSLIDPAESAYAEFLRSKSPQKPSIPIVSSLTGCSLHVMDENFW
NAVKKPMFPEARLYLESQHTCFIDLPGSGTLAAVFKQLIPDGSADRCGSIITPFIHQ
ELKLNMFYFRTPERKFR"
4189..6141
/genes="pkSE"
4189..6141
/genes="pkSE"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13595.1"
/db_xref="GI:2634095"
/db_xref="SPTREMBL:O34787"
/translation="MITYVPGGSGQKMGQGLFEOYQHLTDQAOQILGYSTIEKLCI
EKSLDWNHTEYTPALYVYNALSYLKRVEETGRKPDFAGHSIGEYNALMAAGADF
ETGLRLVLRKGLMGRITGGGMAAVIGLSKEQVTAIVLEEHRLYDIDVANENTPQOIVI
SGPKTEKARAVFENTKQVLFPLNVSGAFHSRYMNEAKQVFKQYIDSFQFAPLAI
PVI SNVYAEYPHODRLKDLTLEQMDNTVKTWDSIRFLMGRGEMEFKAEIGPQVTLGLI
HRIKNEAEPLTYIPKKNPALSALHKEQRNVQAGITAESLGSAAEFKQDYHLTYAYLAGG
MYKGIASKEMVYKLSRAGMGFFGTGLSLKEVEDAIHAIQGELGQAGYGINLVNMM
KHTESEKMDILLNRQVSIVEASFLSVTPVLVRYRAGVKNRQNGDVICSNRLIAK
ISRPEVMSLSPAPENMLQKILGENKITMNEALLRCIPMADDICVADSQSGHTGG
VAYSLPAMTSLDEMMKKYQVRKKIRVGAAGGIGTPEAAMAAFMILGADPILFGTSINO
CTVEAATSDKVKDLQOMNVQDTAYAPAGDMFESGSKVQVLKKGVFEPARANKLYEY
ORVGSIRELDANKLALEEKYFRSV"
6780..8030
/genes="pkSF"
6780..8030
/genes="pkSF"
/function="involved in polyketide synthesis"

```

```

/codon_start=1
/transl_table=11
/protein_id="CAB13596.1"
/db_xref="GI:2634096"
/db_xref="SWISS-PROT:P40804"
/translation="MTKCNLPYVVTVGVTASIGQCKEDFASLSLSGRHAFVDMKRS
GROKDSRFIAEIASLSYDRLSKKMLRKASFSSRAALVTLTEAMBEAELDDADSSRI
GLVVGSGNFQOENFEVYODRSQGFISPAYGLSPMDSLCIGICTDOFGITGLAVTV
GGVASSQOLAVIHALQOVLDSQEDTICIALGALMDLSYMECEALRALGANGTKYADE
PENACRPFQNRDGFYIGESCGALVIERKETALRGLKPYAALSWSIKLUDGNRNDP
SLGEIHVOKALERALLPEDIDYINPHGTGSGFIGDEIELKALRACRLSHAYINATK
SITGHLSAAGIVEIISVLLQKKKSALHPSRNLDHPIDDSFRHVNNEKISISYRIKNALS
LSMGFGMTAVCIQNTIEKGGES"
7946..7974
/genes="pkSF"
8255..9292
/genes="pkSG"
8255..9292
/genes="pkSG"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13597.1"
/db_xref="GI:2634097"
/db_xref="SWISS-PROT:P40830"
/translation="MLITCSESGIDFGKSLSTYIHEVYGLNRCRLPEVYKQACYSCTA
GFOMAYNFILSQTSPGAKALYIASDISRFLIAGGDALEDSDSYAEPGAGAVAVLV
GNEPYVDIPDANGANGYGYEVMOTCRPIDPSEAGSDLSLSYLDCCCEOTFLEYOKRV
PGANYQDTFYIAYHTPPFGMWKVAHRTMMRKAVKVTGSIETDFLTRVYKPLNYCOR
VINGMGAALFLALASTIDQRFDTPKRIGCFYSGSCCSSEFYSGITTPGOERQRTFG
IEKHLDRRYQLSMEYEYELLFKSGMVRFGTRNNVKLDFEMIPGIMQSTQEKPRFLFEEI
SEFHRKRWIS"
9280..10059
/genes="pkSH"
9280..10059
/genes="pkSH"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13598.1"
/db_xref="GI:2634098"
/db_xref="SWISS-PROT:P40805"
/translation="MDLVTVOTIKRFOASVCYITFHRPEANNITINDTLIEECLOVLN
OCTSVTVVVLVLEGLPEVFCFGADFOEIYQEMKRGRKQASSOEPYLDLWMLKLOTGPV
TISHVGRKNVAGLGFVSATDIAIQOTASFSLSSELLFLGLYPACVLPFLIRRRIGROKA
HYMTLTKPISVOEASEWGLDFADESVDVLLRKLRLRLRLLKNGIAHYKQPMSSLD
HOVSRAKALTALTANQDMFSDPQNMGIIRYVETGQFPWEDQ"
10099..10848
/genes="pkSI"
10099..10848
Query Match 100.0%; Score 582; DB 1; Length 233780;
Best Local Similarity 100.0%; Pred. No. 2e-137;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTAATTCCTTATTGATTATTATTTGGCCTACTTGATGACGACATTCATCTGGCTTA 60
Db 150556 ATGTTAATTCCTTATTGATTATTATTTGGCCTACTTGATGACGACATTCATCTGGCTTA 150497
QY 61 ATTGTGGCAAGCTTGCCAAAGGAATTGATATTCGGGAGACGACGAGCGCAACTTAGGC 120
Db 150496 ATTGTGGCAAGCTTGCCAAAGGAATTGATATTCGGGAGACGACGAGCGCAACTTAGGC 150437
QY 121 GCTACCAATGCATCCCGTACATTTGGGTGTAAAAGCTGTTTGGTGGTGCATAGCCGGAGAT 180
Db 150436 GCTACCAATGCATCCCGTACATTTGGGTGTAAAAGCTGTTTGGTGGTGCATAGCCGGAGAT 150377
QY 181 ATTTTGAAGGACACTGGCAACTGCATTCGCTTTTCTCATGATGTTGATATTACCCG 240
Db 150376 ATTTTGAAGGACACTGGCAACTGCATTCGCTTTTCTCATGATGTTGATATTACCCG 150317
QY 241 CTTCTTCAGAGTCTTTGGGTTTGTAGGCCACCTGTTTCCCATCTTCGCCAAATTTAAA 300

```



```
/db_xref="GI:16410546"
/db_xref="SPTREMBL:O8Y7Y9"
/translation="MLTSTGEIAQLFOISKYKIRHYIDEGILVPRKRNPNNGYFFEE
IHSQELMLNEITFVAKODRYEKKVPNOIHDNTINYSOAAKHNLPLDPEEYIFSK
ODGVICRLKRSDDTTPAGEPAGYACKSFIAKDEATIEIOIQLFLEELKYPPTSMILY
ENIYPSLAYPDANVYTMVEVPL"
complement(3288..3293)
/gene="lmo1129"
RBS
complement(3288..3293)
/gene="lmo1129"
RBS
complement(4004..4009)
/gene="lmo1130"
join(4233..4237,4245..5960)
/gene="lmo1131"
4233..4237
/gene="lmo1131"
4245..5960
/gene="lmo1131"
/note="similar to ABC transporters, ATP-binding proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99209.1"
/db_xref="GI:16410547"
/db_xref="SPTREMBL:O8Y7Y8"
/translation="MIDKRLFOLVERKSLVLLILFRVLSLGLMIGLWLVFAQLTHYL
EGONVDMWLGITVVLVGVKAILTKLAEKQIYOASAEELRSMRRVMEKAFRLGNNE
GOLPASLTOLAVDGEIOLEIYYSRFLPOLFYCLIASLMIFGSLVGFAMQPAIVLLIC
MPMPVIMAVMKAKRILSGYSDYNTLGTKEHNSGLSILKAYDQYKXOEIIS
DAERFKATMSLMSQNLNITIMDIISYSGAALGIGMSLIMFTNGTISMTGKMLFLL
SAEFFIPRMLQSLFHVAMNGISACSKLFAYLEKEQVGSVELAKPLEKIEVHHLT
TYEGEKAKLQDYSVHFHNGISFALVKGSGKSTFVRVLLNQLPGYQGEIVWNVPL
ASLSEAIRKQAVDHNHLYANSIRENLLIGNPAANDSDLNWLVPOVSLADFVRKL
PEOLNLEDGNSLGGORQLLALLARLQAEVYVDEITSGVDLESKILLVILQ
ELAKERIVLFIHRLNVLNADQVLFVDFAGKLLLEVASPERLQESYFKNYFLEELAL
LKGA"
5948..5953
/gene="lmo1132"
5948..5953
/gene="lmo1132"
5962..7608
/gene="lmo1132"
5962..7608
/gene="lmo1132"
/note="similar to ABC transporters, ATP-binding proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99210.1"
/db_xref="GI:16410548"
/db_xref="SPTREMBL:O8Y7Y7"
/translation="MSEWTIGWLLKFKPLRGMILAILLGIISLVSIMISLIGTY
GIIVLIQQLPNPKVLFVHVACGVVGRVGLARYLEOYLNHDAIFRLLAIRERIPATLR
KLGARLSGKSGDLVAIITDVEALEVFHAITSPVFIAGTITATVGLASVDVGL
AIIILVQIMVGVVPLMISYKRKKIGTAYQOAFVGLNOMVNIASLODIFOKFGE
ERUKLETQEKLNQYKKRIQSGEQLIGELVGLITATILVILFNGLSOLPLETVLI
GTVLSUSFGVSLNALGALTITTFAGKRLYALTEERKPVILFNGOLELSDFESEAL
DYSFSDHDKOPLNELSLALPKGKWLIGIGESGKSTLVKLMRYWDPDQVKLNN
NELPKITESSLHRLGVMEQRTFLFDLTGNRIIRLGGKDATLEEVKEAARAAIDTWI
ETLPEGYETIIGOSRNLSPGERQIGRLARLFLHDPLLLIDEPSTNLIDEPSTNLID
TLRSEIDQKTVLISHRATTLDAEEQLFENGALKRAAK"
complement(join(7604..7637,7645..7953,7961..7966))
/gene="lmo1133"
complement(7604..7637)
/gene="lmo1133"
7615..7637
/gene="lmo1132"
7615..7637
/gene="lmo1132"
complement(7645..7953)
/gene="lmo1133"
/note="similar to B. subtilis Yjcs protein"
/codon_start=1
/transl_table=11
```

```
/protein_id="CAC99211.1"
/db_xref="GI:16410549"
/db_xref="SPTREMBL:O8Y7Y6"
/translation="MYLKENTNLYCTAVIQTGKVSIEQLVEHLEALRAKTVKEPGC
Query Match 29.5%; Score 171.8; DB 1; Length 250050;
Best Local Similarity 57.8%; Pred. No. 3.5e-33;
Matches 330; Conservative 0; Mismatches 232; Indels 9; Gaps 1;
QY 21 TATTTTGGCTACTTGTATAGGACGATTCATCTGCTTAATTTGGCGAAGCTTGCAA 80
D 147579 TTTATTAGCCTATGTGATTTGCTGATACCATCTGCTTATGGATTCGCAAAATTTCTA 147520
QY 81 AGGAATTTGATATTCGGGACGACGGAAGCGCAACTTAGCGCCTACCACTTCATTCCTGATC 140
D 147519 CAAAAAAGATATTCGGCGATTCGGTAGCGGTAACTTAGGGCGCAAAATTCCTTCCGAGT 147460
QY 141 ATTGGGTGTAAGCTGGTTCGGTCGTCATAGCGGAGATATTTTGAAGGACACATGGC 200
D 147459 GCTTGGTATCAAGCTCGTAGCATTTGTACAGTAATGATATTTTAAAGGAACATGTTGC 147400
QY 201 AACTGCATTTGCTTTTCTCATGCA-----TGTGTATATTCACCCGCTTCTTGCAGG 251
D 147399 AACGTTGCTTCCATCTCTTTCATTTGAATGTAGATCATCATTTCTGTTGCTAACTGG 147340
QY 252 AGCTTTTGGCGTTTTATAGCCACGCTGTTTCCCATCTTCGCCCAAAATTTAAAGCGGTAAAGC 311
D 147339 AGCATTGTCTATTATCGGCGCATAGTTTCCCACTTTTCCGAGGATTCAGAGGTGGGAAAGC 147280
QY 312 CGTGGCGACATCAGGAGCGGCTTTTGTATTTTACGACCCCTGTATTTATATCATCAGATGGT 371
D 147279 AGTAGCTACTTCTCGCGAGGATTCATTCGCTATGCGCCATTACTCTTTGTAGCGGCACT 147220
QY 372 TGCGGTATTTCTCATCTTTTATATCTTACCTTAAATTTGTTCTCTCTCATCATGTTTAAAC 431
D 147219 TGTAGTCTTCTAGTTACACTAAACCTAGTAATATGTATCTCTCAAGTTCAATGATGG 147160
QY 432 AGGCATCTATCTGTTATATATAGTTTCTTTGTCATCATGATCATGATTTATTTGATTTGCTG 491
D 147159 AGCATTGCTGCTAATATTTCTTCTTTTTCATGCGGAGCTGGATTTTAAATTTGCTCGT 147100
QY 492 TACCTGCTCACTATTTTGTGATATACAGACACCGGAGGACATTAAGCAATATATCAA 551
D 147099 TGCTTGTATCGGCTCTTTGTAAATTTGCGCTACCCGAGCAAAATATCACTCGAATACGTA 147040
QY 552 TAAACAGAGACTAAAGTAAATGTTATAA 582
D 147039 CGGAGAGAGCCAAAATAAATGCGATGTA 147009
RESULT 7
AF084044/c
LOCUS
DEFINITION
AF084044 7563 bp DNA linear BCT 07-JUN-2002
Listeria monocytogenes putative LacX gene, partial cds; putative
topoisomerase IV subunit A (parC), and putative LuxS genes,
complete cds; and unknown gene.
ACCESSION
AF084044
VERSION
AF084044.1 GI:21328241
KEYWORDS
Listeria monocytogenes.
SOURCE
Listeria monocytogenes
ORGANISM
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1 (bases 1 to 7563)
AUTHORS
Lampidis, R., Kostrewa, D. and Hof, H.
TITLE
Molecular characterization of the genes encoding DNA gyrase and
topoisomerase IV of Listeria monocytogenes
J. Antimicrob. Chemother. 49 (6), 917-924 (2002)
JOURNAL
MEDLINE
22035523
PUBMED
12039883
REFERENCE
2 (bases 1 to 7563)
AUTHORS
Lampidis, R.
TITLE
Direct Submission
```

JOURNAL Submitted (13-AUG-1998) Institut fuer Medizinische Mikrobiologie und Hygiene, Universitaet Heidelberg, Fakultaet fuer Klinische Medizin Mannheim, Klinikum Mannheim gGmbH, Theodor-Kutzer-Ufer 1-3, Mannheim 68167, Germany

FEATURES Location/Qualifiers

source 1. .7563

 /organism="Listeria monocytogenes"

 /strain="EGD"

 /db_xref="taxon:1639"

 <1..314

 /note="ORFA"

 /codon_start=3

 /transl_table=11

 /product="putative Lack"

 /protein_id="AAM48490.1"

 /db_xref="GI:21328244"

 /translation="GAARYPLNLYDLFKNDALIFEALKOKEMTIKSDKTPHFVKVSFP

 EPFVGWTKAGTTPCLIEPWYGIADGAGESVELRDKAGIEHLEPEAFVASEYIT

 G"

CDS complement(356..952)

 /note="ORFB"

 /codon_start=1

 /transl_table=11

 /product="putative Ynes"

 /protein_id="AAM48493.1"

 /db_xref="GI:21328247"

 /translation="MTINILLSLAYVIGSIPGLWICKFSTKKDIRFEGSGNLGAT

 ISFRVIGRAGSIVTMVILKGTAVTLFFQNLNVDFHFWLTLGAFALIGHISFFLFA

 GRFGKAVATAGVILAYAPLVAFLVFLTKLSKYVLSLSSMITGALAAALIISLFM

 GWLIVLVACALFVIMRHRANITIRNGEFPKIKWM"

 1110..1517

 /note="ORFC"

 /codon_start=1

 /transl_table=11

 /product="putative Yner"

 /protein_id="AAM48491.1"

 /db_xref="GI:21328245"

 /translation="MONPTKEEIOHLKKAKVIYVGLSDKPDRTSYRYSOIMQOAGY

 KIIPNPNADIELGKTVKHLTDIEHVDIVNVFRSEFLPEIAEAFDQIDADVFAQ

 IGIANQAEFLTEKGYPTVMDICIKVATQEMD"

 1698..3665

 /ene="par"

 1698..3665

 /ene="par"

 /codon_start=1

 /transl_table=11

 /product="DNA topoisomerase IV subunit B"

 /protein_id="AAM48488.1"

 /db_xref="GI:21328242"

 /translation="MNRNEYNDISIOVLEGLAVRKRPAMYIGSTDVRLHLVVEIV

 DMSVDTLAGFGKKIVYTLHEDGSVSDGEGMPVGMHKTGKSTVEVILTVLHAGGK

 FOEGGKTSGLHGVGSVYNALSEWLVYTNRDGATYQOKFKDGRKPDGTLKKIGK

 SKATGTTIRFKDPALFPPTSYNFTLSERLRESALFLKGLMLTELDERVGMAEAFH

 EGSKNFVEYINEGDVLHPVASFEGENATIEVEMAFOPNDGYSENILSFVNVRTRG

 AHESGEMKAATRIENDYARVRLNKKDKNDGSDIREGLSVALSIRVPEKILQFE

 GOTKELGTQEAROAVDAVVAEHAYFLAENPETSLLVKKAVKAREAREAKAREE

 TRNGKKRSEVLLSKLPAQSRPNKHELIVLVEGDSAGGSAKQKRRRFOALPLR

 GVYINTERAKDLKANEISITIIHTVGAGVTEFDEDCNDKVVIMDADTDGAHI

 QVLLTFYFVWRPLVAGKVFIAPPLYKVSRSKGEVIEYAMTDEBLSAIOKIG

 KGMIOYRKLGMENADQLMETMNPDTRTLIRVYRVDSSARAERRVATLMDGKVEPR

 QWTEKHVEFSMEDSQNILENNENMVEAKDI"

 3662..6121

 /ene="par"

 3662..6121

 /ene="par"

 /codon_start=1

 /transl_table=11

 /product="DNA topoisomerase IV subunit A"

 /protein_id="AAM48489.1"

 /db_xref="GI:21328243"

 /translation="MSNPEOHIODLALEEVMGDRFGRYSKYIIQERALPDVDRGLKPV

 QRILFAMVEGNTAEKFRKSAKTVGNVIGNYHPHGDSSVYEAIVRMQMDKVRNML

 IEMHNGNSVDGPPAARMRYTEARLSPISAEALLRDIEKETVDFIPNFDOTSSSEPTVLP

ARFPNLVNGSTGISAGYATDIPPHNLTEIEAVIKRLDNPLCTTDDIMKIVKGPDPF

TGGIIOGIDGIRKAYOTGKRVVRSKTEIEDIRGRKOITIEIPYEVNKANLVKRM

DELREKIEGISEVRDETDRGLAVELKDKANAEGVLYLVFKNTDLOVSYNEVMV

AINKRPELMGIIPMLDAYIEHQKEIITIKRSEYDIRKAROHILEGLIKALSILDEV

IKLIRGSKMDRAKLNLTQYDFSEKQAEIVSLQIRUTNTDIEHELSEAKSLAEQI

SVLEKILGDAEELIAVLKEELAEIKKRYKTARRTEVQAEIAEIKIDTEVLVANEIVV

SVTKGYVYKRTORSYAAASGAELAMKEADHAFIOKMNSLDSLLFTSKGMFIYRPV

HELPIDRWNLDGHDVSHLASDLSAGEIRSATAIOAFTKEKRFVTKNGMTPKQSAIT

NYKPORYSKSMMAIKLKGDELLSVHLIDGTEDIPLATKNGYGLVSYITEIPESGART

AGVKAINLKQDDIYIGGIVLMPNEQKHILLATORGSLKQMKASEPEPISRAKRGLLML

RELKSNPHRFIGITLADNNDHLFIETNTDQIIVEIDVANLRITDRYSNNGSVFLDETMEG

EPTSIWLAIEPIKDTSAKEE"

6204..6671

 /note="ORFD"

 /codon_start=1

 /transl_table=11

 /product="putative Luxs"

 /protein_id="AAM48492.1"

 /db_xref="GI:21328246"

 /translation="MAEKMNVESFNLHTKVKAPFVRLAGTKVGVHGDIEIKYDYVRFK

 QPNKEHMEMPALHLEHLMAELARNHTDKLVISPWCOTGYFVSFINHSDYDDALEI

 IATTLTDLVLPATEVPACNEVOGNAASHLEGAKALAEFLAKRSEKRVNPOE"

 7001..57563

 /note="ORFE"

 /codon_start=1

 /transl_table=11

 /product="unknown"

 /protein_id="AAM48494.1"

 /db_xref="GI:21328248"

 /translation="MMSIKSNIMKICVSMVVLPSLTSPLSFAAEETALKESODVV

 NIPDPQLKELNLYNOAENADITEAOAMFRTNITNSGIKDITGIEVLYKNITLISIN

 NINASYEPIQTLSSLEKLVINGENVNADIIPDLSVLSNLTFLDLSRTNIDDTIFSKIS

 NIPNLKLDISNNKIGITKISELNNLPS"

BASE COUNT 2638 a 1298 c 1575 g 2052 t

ORIGIN

Query Match 29.0%; Score 168.6; DB 1; Length 7563;

Best Local Similarity 57.4%; Pred. No. 2.5e-32;

Matches 328; Conservative 0; Mismatches 234; Indels 9; Gaps 1;

QY 21 TATTTTGGCTACTTGATAGCGACGATTCATCTGGCTTAATTTGTGGCGAAGCTTCCCAA 80

DB 926 TTTATTAGCCTATGCTGATTGGTTCGATACCATCTGCTTTATGATGCGAAATTTTCTAC 867

QY 81 AGCAATTGATATTTCGGGACGACGACGACGACGACGACGACGACGACGACGACGACG 140

DB 866 AAAAAAGATATTTCGGGAGTTCGGTAGCGGTAACTTAGGGCGGACAAATTTCTTCCGAGT 807

QY 141 ATTGGGTGTAAGAGCTGGTTCGCTGCTAGCGGAGATATTTTGAAGAGGACACTGGC 200

DB 806 GCTTGTATCAAGCGTGGTAGCATTTGTACAGTAATGGTCAATTTTAAAGGAACACTGTGC 747

QY 201 AACTGCATTGCCCTTTTCTCATGCA-----TGTGTATATTTCACCCCGCTTCTTCAGG 251

DB 746 AACGTTGCTTCCATTTCTTTTTCAGTTGAATGTAGATCATCATTTCTGTTGTTGTAACG 687

QY 252 AGTCCTTTGGGTTTAGGCCACGCTTTTCCCATCTTCGCCAAATTTAAAGGCGGTAAAGC 311

DB 686 AGCAATTGCTATTATTCGGGATAGTTTCCCACTTTTTCGAGGATTCAGAGTGGGGAAGC 627

QY 312 CGTGGCGACATCAGGAGGCTTTTGTATTTTACGACCCCTCTTATTTATCATCAGTGT 371

DB 626 AGTAGCTACTTTCGCCGAGTCAATTTCTGCCTATGGCCCACTTACTCTTTGTACCCGACT 567

QY 372 TCGGCTATTCTTCATCTTTTATATCTTGACTATAAATTTGTTCTCTCTCATCGATTGAAC 431

DB 566 TGTAGCTCTTCTTAGTTACACTAAAACTTAGTAAATATGTATCTCTCAGTTCAATGATTGG 507

QY 432 AGGATCTATACCTTTTATATATAGTTTCTTTGTCATGATACGATATTTTATGATTCGT 491

DB 506 AGCACTTGCTGCAATTAATTTCTCTTTTTCATGGGGAGCTGGATTTTAAATTTCTCTCGT 447

QY 492 TACCTGCTCACTATTTTCTGATATACACACACGCGGACCAACATTAACGAATTAACAA 551

Db 416 TCCTGTATCCGCTCTTTGTAATTTGGCGTCACCGAGCAATATCACTCGAATACGTAA 387
 Oy 552 TAAACAGAACCTAAAGTAAATGTTATATA 582
 Db 386 CGGAGAGAGCCAAATAAATGATGTAA 356

RESULT 8
 LOCUS AE010565 9839 bp DNA linear BCT 25-MAR-2002
 DEFINITION Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 107 of
 197 of the complete genome.
 ACCESSION AE010565 AE009951
 VERSION AE010565.1 GI:19714022
 KEYWORDS
 SOURCE
 ORGANISM

Bacteria; Fusobacteria; Fusobacterium.
 Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T.,
 Lykdis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G.,
 Zhu, L., Vasileva, O., Chu, L., Kogan, Y., Chaga, O., Goltzman, E.,
 Bernal, A., Larsen, N., D'Souza, M., Malunas, T., Pusch, G.,
 Haselkorn, R., Fongstein, M., Kyrpides, N. and Overbeek, R.
 Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586
 J. Bacteriol. 184 (7), 2005-2018 (2002)
 21886394
 11889109

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

2 (bases 1 to 9839)
 Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T.,
 Lykdis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G.,
 Zhu, L., Chu, L., Kogan, Y., Chaga, O., Goltzman, E., Bernal, A.,
 Larsen, N., D'Souza, M., Malunas, T., Pusch, G.D., Haselkorn, R.,
 Fongstein, M., Kyrpides, N. and Overbeek, R.
 Direct Submission
 Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park
 Drive, Chicago, IL 60612, USA

FEATURES
 source

1. 9839
 Location/Qualifiers
 /organism="Fusobacterium nucleatum subsp. nucleatum ATCC
 25586"
 /strain="ATCC 25586"
 /sub_species="nucleatum"
 /db_xref="ATCC:25586"
 /db_xref="taxon:190304"
 232..447
 /gene="FN0528"
 232..447
 /gene="FN0528"
 /codon_start=1
 /transl_table=11
 /product="cold shock protein"
 /protein_id="AAL94724.1"
 /db_xref="GI:19714023"
 /translation="MEVLQMKGTQVFNKKEGFGITCEDGKDVFAHFSQIOKEGKE
 LFEGQVEFEITEQKGQKQASNIWIK"
 complement(619..2100)
 /gene="FN0529"
 complement(619..2100)
 /gene="FN0529"
 /codon_start=1
 /transl_table=11
 /product="Transposase"
 /protein_id="AAL94725.1"
 /db_xref="GI:19714024"
 /translation="MKIPINNKKYFKFPKPLFYINNDIDNDPVRLLSAILSEMDFS
 NLQVFPKTKVFNMFVAYISQKYSKSTDRIEFLCRDSQRTQYLLNSLVNPSYS
 TISFLSKASDIIVELFCQFVKFLSEIPTETIYIDGTFIEAYANKYSFVKKSTL
 KYEKLLENLEIDENKYNKKELDNIDIFSLKLLKIQIKYGRKNSKEQLF
 LEKAQSIVENKNTNLYLEILGERNSFKDKKEATFMKMKEDYMRNQLAPGYNLQIG
 VISEYIASYDFHNPADTKTLIPPLEKTKSONIEIKNVADAGYESFPNVEYLEKNY

VSYKPIYEEKSKTRKYQKNLRNENLEYDEKENRFRKDGLELELEPOYYGDEGKIYF
 KNPETKIIKYNNEFRLLSKSKDNIESDLGKOLRMNRSIQVEGAPAVLKEDMKLRKL
 KYRGKNSRKREIGLCFIAYNFNKYLAKLSRKKQGVVHLPLKTA"
 /gene="FN0530"
 2056..2136
 /gene="FN0530"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAL94726.1"
 /db_xref="GI:19714025"
 /translation="MKKFIYIINWFYHKYIILNEENP"
 2170..2274
 /gene="FN0531"
 2170..2274
 /gene="FN0531"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAL94727.1"
 /db_xref="GI:19714026"
 /translation="MTLNKKKMFVSIFLLGLNLYLCNSPLPFKQIEF"
 complement(2255..3310)
 /gene="FN0532"
 complement(2255..3310)
 /gene="FN0532"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical membrane-spanning protein"
 /protein_id="AAL94728.1"
 /db_xref="GI:19714027"
 /translation="MDIINLVIIAILGGYLADKKVPAAYMLGALFLVAIFNVLF
 NRAFLPNYKEVTOIATGFIGSKPSEDIKMLKVIPGTMVVMVIMAFSFTLSVLM
 STFLGIDNLTSPFATAPCGIMDISLIAYDFKANTSOVALLQIRLISVIFVPPFAKK
 CYERNKKNTSFEREINKKEEKENKESFLFTVIGIIGGIIFSHLPAGTMM
 SCSMALVAFNVKTHKAYMPLTKRIIQSGFAGALIGAKVTLSDVIAKLNLIPIIILII
 IGCLNLIIVGFFLYKTKTFSALLSASPGMSDISLMAEDIGANGPOQVASMQLR
 AIFVIGYPIIILKILFA"
 complement(3332..4372)
 /gene="FN0533"
 complement(3332..4372)
 /gene="FN0533"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical membrane-spanning protein"
 /protein_id="AAL94729.1"
 /db_xref="GI:19714028"
 /translation="MNNKLYGIILCFLLALPAWKLKFFPLVGGPVFGIIGIVIAIL
 LKNRAKFGSINFAKKVLOYAYVILLGFLNLQTIISVSSSLPIIIVSTLSIIAY
 ILAKLINPTKIVILIGVSSICGSAIATAPVINAHDEIAQAISVIFLNFVIAAL
 IPTLGDILNFSNKGFPALFAGTAVNDTSSVTASAWDSHTGTOLDSATVTKLTR
 TLAIIPITLPLAVYNSKRSSNNVNSLKKIFPMFIVYFILASIIITTCVNFIEVGVI
 ENISINNVFSFFKHLKRAFFIIMAVAGLNTNKKLILSGAKPLTLGFCCHFAISL
 VSIQLQKILGIF"
 4639..5067
 /gene="FN0534"
 4639..5067
 /gene="FN0534"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAL94730.1"
 /db_xref="GI:19714029"
 /translation="MNIYVLLARETALYKLFPTWVIYSLPDGLWLFSTGAFLI
 ARKYLHFWFLFWFIYLVWVGIEIQKFGGCHGTPIGTDKTDIIATYATIIIIIA
 LILRKFDNKYKDKTSKEVMQNIYTLIFSVLGLLPNMF"
 5132..5710
 /gene="FN0535"
 5132..5710
 /gene="FN0535"
 /codon_start=1
 /transl_table=11

Qy	61	ATTGTCGGCAGACTTCCCAAGAAATTGATATTCGGGAGCACGGAAGCGCACTTAGGC	120
Db	7424	TGGATAGGAAAAGCTTTTAAAGGTCGTATGTAAGAGATTTATGGAAGTAAATATAGTGGT	7365
Qy	121	GCTACCAATGCATTCCTGTCACATTTGGGTGTAAAGCTGCTCGGTCGCATAGCCGCGAGAT	180
Db	7364	GCTACAAACTCATATAGAGCTTTTAGGAGCAAAATTTGGCGTGTAGCTGTTTTAATAATCGAT	7305
Qy	181	ATTTTGAAAGGACACTGCCAACTGCATTGCCCTTCTTCATGCAAT--GTTGATATTTCAC	237
Db	7304	GTCCTAAAAGGTTTTATACCTCTTTATATGCCAGTAAATTAATTTAGTATATATAATGAT	7245
Qy	238	CCGCTTCTTGCAGGAGCTCTTTCGGTGTTAGGCCACGTGTTTCCCATCTTCGCCAAATTT	297
Db	7244	TTAGTATTATTTAGGTTTACTTCCAAATTTAGCTCATACATTTTCTGCTTTATATCTTTT	7185
Qy	298	AAAAGCGGTAAAGCGGTGGCGACATCAGAGAGCGTTTTCGTAATTTTACGCCCCCTGTA	357
Db	7184	AAAAGTGGAAAAGGTCTTGCAACTAGTTTAGGAGTGTTCCTTATTTTAAATACCTGTTTATA	7125
Qy	358	TTTATCACGATGGTTCGGGTATCTTCATCTTTTATACTTTGACTAAATTTGTTTCTCTC	417
Db	7124	ACTTTAATATATTAGCAATTTTATTTTAGTGGCTTATTTTACTAAATATGTTTTCATG	7065
Qy	418	TCATCATGTTTAAACGGGATCTATACGTGTATATATAGTTCCTTTGTGCCAAT--GATACG	474
Db	7064	GCACTATAACAGCAGCTTTTTTATTGGCAAATTTTACTTTTTTTCACCTATAAAGATTTCA	7005
Qy	475	TATTTATTGATTCGTTACCTCGCTCACTATTTTGTGTATATACAGACACCGAGCGCAAC	534
Db	7004	TATTTATTTCGCTTATACGTTTAAATAGCAGTTTTTTGTTTATATATAGACATATAGACAAT	6945
Qy	535	ATTAAACGAATATCAATAAAAACAGAACCTTAAAGTAAATGGTTATA	581
Db	6944	ATTTCAGATTTACTTGTGCGAACAGAAATAAATTCAAATTTTAAATA	6898

RESULT 9	
AB028896/c	
LOCUS	25021 bp DNA linear BCT 15-AUG-2001
DEFINITION	Streptococcus agalactiae type-specific capsular polysaccharide biosynthesis gene region (cpsIA5 to bcat gene), complete and partial cds.
ACCESSION	AB028896
VERSION	AB028896.2 GI:15186709
KEYWORDS	Uracil DNA glycosylase; ung; orf1; NeuA; NeuD; NeuE; NeuF; NeuG; NeuH; NeuI; NeuJ; NeuK; NeuL; NeuM; NeuN; NeuO; NeuP; NeuQ; NeuR; NeuS; NeuT; NeuU; NeuV; NeuW; NeuX; NeuY; NeuZ; NeuA1; NeuA2; NeuA3; NeuA4; NeuA5; NeuA6; NeuA7; NeuA8; NeuA9; NeuA10; NeuA11; NeuA12; NeuA13; NeuA14; NeuA15; NeuA16; NeuA17; NeuA18; NeuA19; NeuA20; NeuA21; NeuA22; NeuA23; NeuA24; NeuA25; NeuA26; NeuA27; NeuA28; NeuA29; NeuA30; NeuA31; NeuA32; NeuA33; NeuA34; NeuA35; NeuA36; NeuA37; NeuA38; NeuA39; NeuA40; NeuA41; NeuA42; NeuA43; NeuA44; NeuA45; NeuA46; NeuA47; NeuA48; NeuA49; NeuA50; NeuA51; NeuA52; NeuA53; NeuA54; NeuA55; NeuA56; NeuA57; NeuA58; NeuA59; NeuA60; NeuA61; NeuA62; NeuA63; NeuA64; NeuA65; NeuA66; NeuA67; NeuA68; NeuA69; NeuA70; NeuA71; NeuA72; NeuA73; NeuA74; NeuA75; NeuA76; NeuA77; NeuA78; NeuA79; NeuA80; NeuA81; NeuA82; NeuA83; NeuA84; NeuA85; NeuA86; NeuA87; NeuA88; NeuA89; NeuA90; NeuA91; NeuA92; NeuA93; NeuA94; NeuA95; NeuA96; NeuA97; NeuA98; NeuA99; NeuA100; NeuA101; NeuA102; NeuA103; NeuA104; NeuA105; NeuA106; NeuA107; NeuA108; NeuA109; NeuA110; NeuA111; NeuA112; NeuA113; NeuA114; NeuA115; NeuA116; NeuA117; NeuA118; NeuA119; NeuA120; NeuA121; NeuA122; NeuA123; NeuA124; NeuA125; NeuA126; NeuA127; NeuA128; NeuA129; NeuA130; NeuA131; NeuA132; NeuA133; NeuA134; NeuA135; NeuA136; NeuA137; NeuA138; NeuA139; NeuA140; NeuA141; NeuA142; NeuA143; NeuA144; NeuA145; NeuA146; NeuA147; NeuA148; NeuA149; NeuA150; NeuA151; NeuA152; NeuA153; NeuA154; NeuA155; NeuA156; NeuA157; NeuA158; NeuA159; NeuA160; NeuA161; NeuA162; NeuA163; NeuA164; NeuA165; NeuA166; NeuA167; NeuA168; NeuA169; NeuA170; NeuA171; NeuA172; NeuA173; NeuA174; NeuA175; NeuA176; NeuA177; NeuA178; NeuA179; NeuA180; NeuA181; NeuA182; NeuA183; NeuA184; NeuA185; NeuA186; NeuA187; NeuA188; NeuA189; NeuA190; NeuA191; NeuA192; NeuA193; NeuA194; NeuA195; NeuA196; NeuA197; NeuA198; NeuA199; NeuA200; NeuA201; NeuA202; NeuA203; NeuA204; NeuA205; NeuA206; NeuA207; NeuA208; NeuA209; NeuA210; NeuA211; NeuA212; NeuA213; NeuA214; NeuA215; NeuA216; NeuA217; NeuA218; NeuA219; NeuA220; NeuA221; NeuA222; NeuA223; NeuA224; NeuA225; NeuA226; NeuA227; NeuA228; NeuA229; NeuA230; NeuA231; NeuA232; NeuA233; NeuA234; NeuA235; NeuA236; NeuA237; NeuA238; NeuA239; NeuA240; NeuA241; NeuA242; NeuA243; NeuA244; NeuA245; NeuA246; NeuA247; NeuA248; NeuA249; NeuA250; NeuA251; NeuA252; NeuA253; NeuA254; NeuA255; NeuA256; NeuA257; NeuA258; NeuA259; NeuA260; NeuA261; NeuA262; NeuA263; NeuA264; NeuA265; NeuA266; NeuA267; NeuA268; NeuA269; NeuA270; NeuA271; NeuA272; NeuA273; NeuA274; NeuA275; NeuA276; NeuA277; NeuA278; NeuA279; NeuA280; NeuA281; NeuA282; NeuA283; NeuA284; NeuA285; NeuA286; NeuA287; NeuA288; NeuA289; NeuA290; NeuA291; NeuA292; NeuA293; NeuA294; NeuA295; NeuA296; NeuA297; NeuA298; NeuA299; NeuA300; NeuA301; NeuA302; NeuA303; NeuA304; NeuA305; NeuA306; NeuA307; NeuA308; NeuA309; NeuA310; NeuA311; NeuA312; NeuA313; NeuA314; NeuA315; NeuA316; NeuA317; NeuA318; NeuA319; NeuA320; NeuA321; NeuA322; NeuA323; NeuA324; NeuA325; NeuA326; NeuA327; NeuA328; NeuA329; NeuA330; NeuA331; NeuA332; NeuA333; NeuA334; NeuA335; NeuA336; NeuA337; NeuA338; NeuA339; NeuA340; NeuA341; NeuA342; NeuA343; NeuA344; NeuA345; NeuA346; NeuA347; NeuA348; NeuA349; NeuA350; NeuA351; NeuA352; NeuA353; NeuA354; NeuA355; NeuA356; NeuA357; NeuA358; NeuA359; NeuA360; NeuA361; NeuA362; NeuA363; NeuA364; NeuA365; NeuA366; NeuA367; NeuA368; NeuA369; NeuA370; NeuA371; NeuA372; NeuA373; NeuA374; NeuA375; NeuA376; NeuA377; NeuA378; NeuA379; NeuA380; NeuA381; NeuA382; NeuA383; NeuA384; NeuA385; NeuA386; NeuA387; NeuA388; NeuA389; NeuA390; NeuA391; NeuA392; NeuA393; NeuA394; NeuA395; NeuA396; NeuA397; NeuA398; NeuA399; NeuA400; NeuA401; NeuA402; NeuA403; NeuA404; NeuA405; NeuA406; NeuA407; NeuA408; NeuA409; NeuA410; NeuA411; NeuA412; NeuA413; NeuA414; NeuA415; NeuA416; NeuA417; NeuA418; NeuA419; NeuA420; NeuA421; NeuA422; NeuA423; NeuA424; NeuA425; NeuA426; NeuA427; NeuA428; NeuA429; NeuA430; NeuA431; NeuA432; NeuA433; NeuA434; NeuA435; NeuA436; NeuA437; NeuA438; NeuA439; NeuA440; NeuA441; NeuA442; NeuA443; NeuA444; NeuA445; NeuA446; NeuA447; NeuA448; NeuA449; NeuA450; NeuA451; NeuA452; NeuA453; NeuA454; NeuA455; NeuA456; NeuA457; NeuA458; NeuA459; NeuA460; NeuA461; NeuA462; NeuA463; NeuA464; NeuA465; NeuA466; NeuA467; NeuA468; NeuA469; NeuA470; NeuA471; NeuA472; NeuA473; NeuA474; NeuA475; NeuA476; NeuA477; NeuA478; NeuA479; NeuA480; NeuA481; NeuA482; NeuA483; NeuA484; NeuA485; NeuA486; NeuA487; NeuA488; NeuA489; NeuA490; NeuA491; NeuA492; NeuA493; NeuA494; NeuA495; NeuA496; NeuA497; NeuA498; NeuA499; NeuA500; NeuA501; NeuA502; NeuA503; NeuA504; NeuA505; NeuA506; NeuA507; NeuA508; NeuA509; NeuA510; NeuA511; NeuA512; NeuA513; NeuA514; NeuA515; NeuA516; NeuA517; NeuA518; NeuA519; NeuA520; NeuA521; NeuA522; NeuA523; NeuA524; NeuA525; NeuA526; NeuA527; NeuA528; NeuA529; NeuA530; NeuA531; NeuA532; NeuA533; NeuA534; NeuA535; NeuA536; NeuA537; NeuA538; NeuA539; NeuA540; NeuA541; NeuA542; NeuA543; NeuA544; NeuA545; NeuA546; NeuA547; NeuA548; NeuA549; NeuA550; NeuA551; NeuA552; NeuA553; NeuA554; NeuA555; NeuA556; NeuA557; NeuA558; NeuA559; NeuA560; NeuA561; NeuA562; NeuA563; NeuA564; NeuA565; NeuA566; NeuA567; NeuA568; NeuA569; NeuA570; NeuA571; NeuA572; NeuA573; NeuA574; NeuA575; NeuA576; NeuA577; NeuA578; NeuA579; NeuA580; NeuA581; NeuA582; NeuA583; NeuA584; NeuA585; NeuA586; NeuA587; NeuA588; NeuA589; NeuA590; NeuA591; NeuA592; NeuA593; NeuA594; NeuA595; NeuA596; NeuA597; NeuA598; NeuA599; NeuA600; NeuA601; NeuA602; NeuA603; NeuA604; NeuA605; NeuA606; NeuA607; NeuA608; NeuA609; NeuA610; NeuA611; NeuA612; NeuA613; NeuA614; NeuA615; NeuA616; NeuA617; NeuA618; NeuA619; NeuA620; NeuA621; NeuA622; NeuA623; NeuA624; NeuA625; NeuA626; NeuA627; NeuA628; NeuA629; NeuA630; NeuA631; NeuA632; NeuA633; NeuA634; NeuA635; NeuA636; NeuA637; NeuA638; NeuA639; NeuA640; NeuA641; NeuA642; NeuA643; NeuA644; NeuA645; NeuA646; NeuA647; NeuA648; NeuA649; NeuA650; NeuA651; NeuA652; NeuA653; NeuA654


```

CDS
/gene="cpsIas"
<1..375
/gene="cpsIas"
/function="putative histidine protein kinase"
/codon_start=1
/transl_table=11
/product="CpsIas"
/protein_id="BAB62877.1"
/db_xref="GI:15186710"
/translation="DHYDTWKFKELKESNHPVLIAFGERHWDRLTSGKSLAEQLQITD
LDEVKSTIQLOREFKSVRNPLAHLKPFDEQELRYTQFSOAFDQIIFLAKVIG
VEYDVNFHDTVNKLIIKILE"
complement(400..1323)
/gene="cpsIar"
complement(400..1323)
/gene="cpsIar"
/function="LysR-type transcriptional regulator"
/codon_start=1
/transl_table=11
/product="CpsIar"
/protein_id="BAB62878.1"
/db_xref="GI:15186711"
/translation="MRIOQLQYVIVKIVETGSMNEAAKOLVITOPSLSNVARNLETEMG
IQFIRNPKGITLTKDMEFLSVARQILEOTALLEERYKGDNTSRELFVSSSHAYEV
VNARVALNGTDMQYELFELREHETWELIIDVNFRSEIGULFNLNTRDVLKLFDD
NSLIATLITFTTHIFVSKSNPLANRKLKMKDLEDPYLSIDGOLKSTIYTYKIKKI
QIPHKTSIVSDRATFLNLMIGLDGYTVATGILNSKLNGDEIVAIPLDVDDVIDIVI
RHDKANLSKMGQKFDYLLLEVSFHNKQKS"
complement(1447..1452)
/feature="misc_feature"
1453..1457
/evidence="experimental"
-10_signal
1458..1482
/evidence="experimental"
-35_signal
1483..1486
complement(1481..1486)
/evidence="experimental"
-10_signal
1487..1512
complement(1507..1512)
/evidence="experimental"
-35_signal
1512..2969
/gene="cpsIaa"
1512..2969
/gene="cpsIaa"
/function="regulator"
/codon_start=1
/transl_table=11
/product="CpsIaa"
/protein_id="BAA82275.1"
/db_xref="GI:5381178"
/translation="MSNHSRRQKKHSHHTPLRVINLFLVIFILLSVYSLFMYRRHF
LAFRLNAVIGVTVLVLILASLFLCIKNKARIEFTIILVLASIFVATTLXGKSTIDL
TNLNKATSYSEIEMSVIVPKDSKITNEAVSKLAAPVKNDTSNITDLIEHISEKGI
SITPKQTSVODANVRNKGDSQAMVNNAYVSLIELSTPDFKSOIKTIYTYKIKKI
NRKNTKEGVFNIIYSGIDTFCGSITSVSRSDVINIMVNTNHHKVLTLTPPDAYVK
IPDGGQYIKLTHAGLVGVETSMKLTENLYDINLDYIRINFSSFLKLDLGGVTV
YNDQAFYSKHGDFPVQVTLNSEQALGFVRERYSLQGGDNRGRQERVIAGIINK
LASSQVTKLNSITQSLOTQSVQNTNMTIDNIDNLINQLSTQRTFVESQALTGHGVTG
ELPSYAMPGAQLYMWSDQSLSNAKSKIKNTWEE"
2975..3706
/gene="cpsIab"
2975..3706
/gene="cpsIab"
/codon_start=1
/transl_table=11
/product="CpsIab"
/protein_id="BAA82276.1"
/db_xref="GI:5381179"
/translation="MIDIHSHIVFDVDDGPKTLEESLSLEESYRQGVRIIVSTSHRR
KGMFETEDIFKFNFSIVKHEAKRFEHLQILYGGELYTSDMLEKLLKLOIPTLNNT
KFALIEFSMOTSKNDIHTALSNLMLGITPVVAHIERYNALENOKERVKEIINNGCVT
QINSHILKOKLFNDKHKRKYRFLNLFVHVASDMMHNDLVRPFLAEAYKIC
RDFGRANQLFIENAOISILKNHYI"

gene
3715..4407
/gene="cpsIac"
3715..4407
/gene="cpsIac"
/function="chain length determination (putative)"
/codon_start=1
/transl_table=11
/product="CpsIac"
/protein_id="BAA82277.1"
/db_xref="GI:5381180"
/translation="MNKIANTVEINFLNLLKLMKKKFLITFVAIAFATAGLFYSILF
IVTPOYISSPRTIYVNPNTPNNSITADQLQAGSFANDYKEIITSDVLEKVISSEKL
NYFSSQLQKITVLSIKDTRVISVEDANPKMSQKLANYSREARVSKIRAKVQVEDI
TTLEKGNLPRAPSPNPNIKNVLGIFVAGLSTIVLVIMGILDDRNVTEEDIEKALGL
TSLGIVPDLNKL"
4418..14922
/feature="misc_feature"
4418..5107
/gene="cpsIad"
4418..5107
/gene="cpsIad"
/function="chain length determination (putative)"
/codon_start=1
/transl_table=11
/product="CpsIad"
/protein_id="BAA82278.1"
/db_xref="GI:5381181"
/translation="MTRLEIVDSKLRQAKKTEEVFNARTNIQFSGKENKILAITSVR
EGEKSSTSTSLASLAOAGFKTLLIDADTRNSVMSGTPKATCTIKGLTNVLSGNADL
GDIICTNVPRLMVPSGKVPNPNTALLONAYPNKKEIAKNIFDIITDTPPIGLVY
DAAIIANACDGFILVTQAGRIKKNRYEKAQEQSGSKFLGILNKNVSESATYGDY
GNYGKRRDKRK"
5243..5248
/evidence="experimental"
-35_signal
5265..5270
/evidence="experimental"
-10_signal
5334..6509
/gene="cpsIae"
5334..6509
/gene="cpsIae"
/function="glucosyltransferase"
/codon_start=1
/evidence="experimental"
/transl_table=11
/product="CpsIae"
/protein_id="BAA82279.1"
/db_xref="GI:5381182"
/translation="MWLKYSFYVIFISSLFFIFPKNSFTTTRLSFFSFTAMNSILLYL
LNSFLATYRKYSYAKFSRDTKVVLTITNKDSLKMTFRNKHYNHYIAVCILDSSEKDY
DLKHSRLRIINKDALTSELTCLTVDAQFINPIELFKYQIOIDIINDIEAMGVIVN
VEALSFDNIGEKRIOTFEGSVITYSMKFYKSHLIAKRLDITGAILIGLICGIVAI
FLVPOIRKDGGAIPFSONRVNRGRIFRFYKFSMRVDAEQIKKDLLVHNMGTGLMFK
LDDDPRTIKGFIKRTSIDELPQFYNVLKGDMSLVCTRPPTVDEYKYNSTOKRRLS
PKGIVGLWQISGRNNITDFDEIVKLDVQYINENWSIWSDIKIILLTLKVVLGTGAK"
6533..6982
/gene="cpsIaf"
6533..6982
/gene="cpsIaf"
/codon_start=1
/transl_table=11
/product="CpsIaf"
/protein_id="BAA82280.1"
/db_xref="GI:5381183"
/translation="MKICLVSGSGHLAHLNLLKLPFGKNKDRFWVTFDKEDARSILRE
EIVYHGFEPPTNRNVKLVNKTILAFKRLKRPDVLISSAAVAVPFFYIGKLGCKT
VYIEVDRIDKPTLTGKLVYPVTDKFIQVQEMKKVYPRAINLGGIF"
6982..7524
/gene="cpsIag"
6982..7524
/gene="cpsIag"
/function="galactosyltransferase"
/codon_start=1
/transl_table=11

```

```
/gene="parE"  
/codon_start=1  
/transl_table=1  
/product="topoisomerase IV"  
/protein_id="CAP40548.1"
```

[illegible]


```
/gene="deod"
/notes="spym18_0955"
362..1075
/gene="deod"
/notes="best blastp match gb|AAK33813.1| (AE006539)
putative purine nucleoside phosphorylase [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative purine nucleoside phosphorylase"
/protein_id="AAL97596.1"
/db_xref="GI:19748121"
/transl_table="MSIHSKAGKDIADKILLPCDPLRAKFAIENFLEDAFCNEVRN
MFGVTGYKGRHVSVMGTGMGSPSISYIAVELJVDYGVKTLIRVGTAGADPVDHVR
LVAQRAATNSIIRNDPREFDPQIADFGLLDKAYHAREMGVTHVGNVLSDFY
TNPTFNMALGKGVRAIENEAALYLAHQHVKALGIMTISDNLNDPTEDTTAER
QTFTDMKVGLETLIAND"
1062..1856
/gene="spym18_0956"
1062..1856
/gene="spym18_0956"
/notes="best blastp match gb|AAK33814.1| (AE006539)
histidine protein kinase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="histidine protein kinase"
/protein_id="AAL97597.1"
/db_xref="GI:19748122"
/transl_table="MPMTNOTLDILDVYVYAHAFRIAKAPNIPKTVLYLLEMLKE
RRELNLAFLAHAENRTIEDYHCSLWLNOSLEDEQIANIYILDLEVKVNGAIIDFV
RSVSPILYFLRLITSEIYFNKAYIFDKNDQDVTWHFOAMLESDEHVEFKAYLSOKO
SRNVTTSKSLADMLTSLPQELDKVFLPFLHFEKAVENPLAHLIKPDEEELHRTTHF
SSQAFLENIITLAFSGVIYRPEPFYDDMNAIKKELSLWRQSV"
/complement(1936..2841)
/gene="cpsy"
/notes="spym18_0957"
complement(1936..2841)
/gene="cpsy"
/notes="best blastp match gb|AAK33815.1| (AE006539)
putative transcriptional regulator [Streptococcus pyogenes
M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL97598.1"
/db_xref="GI:19748123"
/transl_table="MRIOQLHYIKIIVCGSMNEAAKOLFITOPFSUNAVKDLMEMNG
ITIFNRNPKGITLTKDGEFLSVARQIIQTSLEDRYKNHNTGRELFVSYSQYAFV
VNAFVSLKRTDTRYELFRETRETWEIIDQVKNFRSEIGLVLEINDYNRDLTKLFD
NHLTASPLFRAQPHIFVYSKNPLATKSLSDMDLDRDPYLSYDQGHNSFYFSEEMMS
OMPHNKSIVVSDRATFLNMLIGDGTVASGILNSNLNGDQIIVAIPLDVPDEIDVFI
KHEKANLSKGRERFYLEEVFD"
3063..3599
/gene="spym18_0958"
3063..3599
/gene="spym18_0958"
/notes="best blastp match gb|AAK33816.1| (AE006539)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97599.1"
/db_xref="GI:19748124"
/transl_table="MNVTEQNSHOILOKLLVSIHYITLFRDELKLVERTPSILGGFF
PAHLVQSELDGIVAAIDTLDQQRQLESTFWEESAFKLANKTLDIDVNNIKGVYHLI
DLQSQKEVFOIIGDKRIRVFGVLSDFSSLKVSALSLEKAPIDVLYEHIQRYNLEE
EATKYQSPKVPPTGEF"
3876..4568
/gene="pyrF"
/notes="spym18_0959"
3876..4568
/gene="pyrF"

/notes="best blastp match gb|AAK33817.1| (AE006539)
putative orotidine-5'-decarboxylase PyrF [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative orotidine-5'-decarboxylase"
/protein_id="AAL97600.1"
/db_xref="GI:19748125"
/transl_table="MKEERPIALDFSSPEETKAFLDLPAAEEKLYVKIGMELYAQQ
PDIVRIKSLGHNVFLDLKHLIDPNTVRAMAVLKELDIDMATVHAAGVEMUKAARE
GUGQDPTLIATVOLTSTSEDMRGDQNIQTSLSLESYHSGAAKALDQGVVCSAQEV
EAIKAYTPTGFTCTLPTGIRPKGSIQDQKRVMTPNQARRIESDYIVVGRPIITQAKDPV
AAQQAIAKAWAG"
4626..5255
/gene="pyrE"
/notes="spym18_0960"
4626..5255
/gene="pyrE"
/notes="best blastp match gb|AAK33818.1| (AE006539)
putative orotate phosphoribosyltransferase [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative orotate phosphoribosyltransferase"
/protein_id="AAL97601.1"
/db_xref="GI:19748126"
/transl_table="MTLASQIATQLDIDKAVYLKPEDPFTWASGSIKSPIYTDNRVTL
YKPTROLDIEGVETIRAHFVEVEIAGTAGATAGPHGAIADKMTLPFAYIRSKPKDH
GAGNOIEGVLKQKMWIIEDLSTGGSVLDAASAAAGADVGLVITFYIYELPKAS
ONFKEAGIKLITLSNVTETLIAVAKLQYITNDGLHLKFKEDQVNNQQ"
5452..6906
/gene="amic"
/notes="spym18_0961"
5452..6906
/gene="amic"
/notes="best blastp match gb|AAK33819.1| (AE006539)
putative amidase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="AAL97602.1"
/db_xref="GI:19748127"
/transl_table="MTYQATAMAIQVOTGOTPLELVTOAIYKAKKLNPLNATISE
RFEALAEAKORDPFSGLPFAGVPIFLKDLGOELKHSSTSGSRLFKFYQATKTDLFVK
RLAALGFILGRSNTPEFGKNSLSDSLHCPVNLKPDNTRNAGSGSGGAALVSSGIS
ALAPASGGGSIKIPASFNGLICKSRGMPVPGPSYRSWQSGASVHFALTYSVROTR
NLLYIQOMESPPFLATLTKDSIYQSQRPLTIAFYORLSGSPVSLDTAKALQOA
VWLRQGHQVLEEFVNMTEVIRHYIMNSVETAAFMADIEDTLGRMPTKDDMET
MTWAIYQSGKDIIPAWRYSOVLQKWDYTSATMASFHYETYNLLTFTTNTAPKRGELVP
DSKLMANLQAQEIFSSEEDFNLVETMFGKSLAINPYTALPNLTGQPAISLPTVETREG
LPMGIQLIAAKGREDLLLGIAEQFEAAGLLKIPQ"
7024..7887
/gene="spym18_0962"
7024..7887
/gene="spym18_0962"
/notes="best blastp match gb|AAK33820.1| (AE006539)
putative ABC transporter (binding protein) [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ABC transporter (binding protein)"
/protein_id="AAL97603.1"
/db_xref="GI:19748128"
/transl_table="MKKSVTLISIGLASLLAACAPHKSKQSSWNTIKEGVLKVATTP
GTYOPTSFYNDNNELGYEVDYMKKEIGKRLNIKRVKFEVTFDQAFVSDVSGRYDISL
NFDTIPKRRQYINISYPIKYGVGMIVRADGSSNMMAKDLSDMKGKGAAGSCTEYMK
VAQOQAEWYTDYNTGVYLVNDVANGRTDFIPNDYPAOKLFYDYLNSOLPNLNVKMS
DVQYNPTGEGIVMKNKDDSLKXIDDAVIDMKKDGSLKRISETIYAGODLTKEPFGDK
KIPVIDTKDN"
7916..8566
/gene="spym18_0963"
7916..8566
/gene="spym18_0963"
```

```

/note="best blastp match gb|AAK33821.1| (AE006539)
putative ABC transporter (permease) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (permease)"
/protein_id="AAL97604.1"
/db_xref="GI:19748129"
/translators="MINIPLMKDSLGFLVSLPVTIGISLLSFLTGIVFGLLALLGR
SROPLHLYRTYISINRGYPHIVLVFLYGLPYGLLEPALLCAVLGRSMVSAAYI
SEVRSIADIGQWEAKALGPLYALMKKILPOAFRIAPVPLGNVIDMWKSS
LAAMITVPDIFQNAKITGGREWDYMSYIIVAFIYLIWFLIERYOEFLENKALV"
8699..9352
gene
/gene="ung"
/note="spym18_0964"
8699..9352
CDS
/gene="ung"
/note="best blastp match gb|AAK33822.1| (AE006539)

Query Match          19.6%; Score 114; DB 1; Length 11573;
Best Local Similarity 53.7%; Pred. No. 1.8e-18;
Matches 259; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Oy 1  ATGTTAATGCTTTATGATATTATTTGGCCCTACTGATAGCAGCATTCCTATCGGCTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11478 ATGAATTAATACTATTTTATTACCAATTCCTATTACTTGTCTATTATCCAACTGGACTA 11419

Oy 61  ATTGGGCAAGCTTCCCAAGGAATTCATATTGGGAGCAGGAAGCGGCAACTTAGGC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11418 TGGATTGGACAGTACTTTTACCACATCACTTACGAGAGCATGGATCAGGAATACTGGA 11359

Oy 121  GCTACCAATGATTCCTCGTACATTTGGTGTAAAGCTGCTTCGGTCTCATAGCCGGAGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11358 ACCACAAATACTTTTCGATTTTAGGTGTCAAGCAGCAAGCATCTTAGCTATTGAT 11299

Oy 181  ATTTTGAAGGACACTGGCAACTGCATTTGCTTTTCTCATGCAATG---TCATATTCCAC 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11298 ATGTTTAAAGGACACTTTCAATATTGTTACCAATTTATTTTGGTATGACTTCAATTCA 11239

Oy 238  CCGCTTCTTGAGGAGTCTTTGCGGTTTTAGGCACAGTGTTCCTCATCTTCGCCAAATTT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11238 TCCATTGCTATCGCGTTTTTCGCAGTTTTTAGGCATACTTTTTCCTATTATTTTGCCTCAACTTT 11179

Oy 298  AAAGCGCTAAAGCGTGGCGCATCATCAGGAGCGTTTTGCTATTATTACGACACCCTGTTA 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11178 AAAGTGTGTAAGCGCTAGCAACAAGTCTGCTGTATTGCTAGGCTTTCCTCGCTTATAT 11119

Oy 358  TTTATCAGATGGTTCGGGTATTCTTCATCTTTTATATCTTGAATAATTTGTTCTCTC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11118 CTCTTTTTTTAGCATCTATCTTTGTTTCTCTATTTTATTAGTATCATATCTTTTA 11059

Oy 418  TCATCGATGTTAACAGGATCTATACCTTTATATATAGTTTCTTTGTCCTATGATCGTAT 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11058 GCTAGTGTGGTTTCAGCATCTCGTTGGTGTGTATCTGTTTAAACATTTCTCGCCATCAT 10999

Oy 478  TT 479
    ||
Db 10998 TT 10997

RESULT 13
AE014149/c
LOCUS
DEFINITION Streptococcus pyogenes MGAS315, section 14 of 37 of the complete
genome.
ACCESSION AE014149
VERSION AE014149.1
KEYWORDS GI:21904329
SOURCE Streptococcus pyogenes MGAS315.
ORGANISM Streptococcus pyogenes MGAS315.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 53354)

```

```

AUTHORS Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
TITLE Genome sequence of a serotype M3 strain of group A Streptococcus:
Phage-encoded toxins, the high-virulence phenotype, and clone
emergence
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
REFERENCE 2 (bases 1 to 53354)
AUTHORS Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
FEATURES
source
1..53354
/organism="Streptococcus pyogenes MGAS315"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/notes="group: A"
123..986
/gene="thya"
/note="Spym3_0601"
123..986
/gene="thya"
/notes="best blastp hit: gb|AAK33802.1| (AE006538) putative
thymidylate synthase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative thymidylate synthase"
/protein_id="AAM79208.1"
/db_xref="GI:21904330"
/translators="MLTERLEMTKADQIFKANTOKIINEGSLSEQARPKYKDGRTAH
SKYTCGAEDYLAKEFPITTLRPIKSAIWELEWYODOSNLDVLEAKYNVHW
NEVEDOTRTIGORYGAVKHDIISIKLAELAEWPNRRNVLISMDVFEETKGLL
PCAFQIMFDRVRGEDLYLDAISRNSDILVAHHINAMQYVALOMHIAKHFQKIGK
FFYFVNLIHYDNQFOAQELLKRPVQSKLVLYNPVRTNFFDINKPDDFELQNVDP
VKQLHFDLAI"
1066..1563
/gene="dys"
/note="Spym3_0602"
1066..1563
/gene="dys"
/notes="best blastp hit: gb|AAK33803.1| (AE006538) putative
dihydrofolate reductase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative dihydrofolate reductase"
/protein_id="AAM79209.1"
/db_xref="GI:21904331"
/translators="MTKEIILIAWAEAGLIGVAGLPWYLPKLEHFKKTLHQAIL
MGRVTFEGMNCRLPOROTLVMTNRDYOVDVLTMTSIEKLVLEWHAODKTLVIGG
NKVLEAFNGYFDRILIKTVIHREFKGTYPRLNDFSHTQESQFEYADAKNPDTFTVT
ALXHK"
1583..1753
/gene="Spym3_0603"
1583..1753
/gene="Spym3_0603"
/notes="best non-GAS blastp hit: gb|AAK75657.1| (AE007451)
hypothetical protein [Streptococcus pneumoniae TIGR4], and
gb|AAL00232.1| (AE008512) hypothetical protein
[Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM79210.1"
/db_xref="GI:21904332"
/translators="MORSIFGVFTAFCLVICVLCATPAFKKNRYGLASFLNAFTNL
VNSHAFYGTLF"

```

```

1883. 3112
/gene="clpX"
/note="Spym3_0604"
1883. 3112
/gene="clpX"
/note="best blastp hit: gb|AAK33805.1| (AE006538) putative
ATP-dependent Clp protease subunit x [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent Clp protease subunit x"
/protein_id="AAW79211.1"
/db_xref="GI:21904333"
/translation="MAGSRNDIKVYCSFCGSKQDDVKKIAGNNVFCNECVALSOE
IRKEELAEVLADTEVPKRELDVLNQYVGGDRAKRALSVAVNHYKRVSTESR
DDDDVDLQKNIIMIGPTSGKTFLAQTLAKSLNPPFAIADATSLTEAGVGEDVENI
LLKLIADYINRAERGIIVDEIDIAKGNVSITRDVSGEGVQOALLIEGTIV
ASVPQGRKHPNEMIQIDTKNIFIVGGAFDGEIEIVKORLGEKIVGFQNSRKID
DNASYMOELISEDLQKLELPEFIRGLPVVALEOLNTSLIOILLTEPRNALVKYQA
LLSDGVELAFDRALEAIANKAIAERTGANGRLSIIEETWLDIMFEIPSOEDVTKVR
ITRAAVEGSKPVLETA"
3122. 3721
/gene="Spym3_0605"
3122. 3721
/gene="Spym3_0605"
/note="best blastp hit: gb|AAK33806.1| (AE006538)
conserved hypothetical GTP-binding protein [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein"
/protein_id="AAW79212.1"
/db_xref="GI:21904334"
/translation="MAEEQVLNTHNASILLASANKSHYPQDDLPETALAGRSNVGKSS
FINILKKNLARTSSKPGTKQLNFNIDDKLRFVDVPGYAKVSKSRKWKRM
EYLTSRDNLRAVSLVDLRHAPSKEDIQMDYFLKYDIPVIVVATKADKIPHGKWK
HESVVKALNPFKSDTFIVFSVERIGIDDSWDAILEQV"
3869. 4615
/gene="Spym3_0606"
3869. 4615
/gene="Spym3_0606"
/note="best blastp hit: gb|AAK33807.1| (AE006538)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW79213.1"
/db_xref="GI:21904335"
/translation="MTOVAOGRSKGKSLCHOCNALFVFIIFILLAVLFGVRYRNS
LEGWRTSIDOKLGDFAKRLTGLHOSPLIDDSLLSSOMILTKNNVNDLSEVQV
ERDIFARLAAHYONELLTKLKNHLVVGDLSSKEROIENSMPASHELEMLDQAFE
KLASQIGGKYNQKTHLSAVVVGKVNRIHTIDIEEVAAGHTSFSKGLLTPNGYED
YTRFGKLELIGDEKIIIFKALKKSVKCI"
complement(4668. 6767)
/gene="clpI"
/note="Spym3_0607"
complement(4668. 6767)
/gene="clpI"
/note="best blastp hit: gb|AAK33808.1| (AE006538) putative
ATP-dependent Clp proteinase (ATP-binding subunit)
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent Clp proteinase
(ATP-binding subunit)"
/protein_id="AAW79214.1"
/db_xref="GI:21904336"
/translation="MTHFSKDPFVNMDIDFNOLMANMGVRSNPRLVNGREITPE
EFQIIRQTGQPVATTKATNSQMLTPKADSVLTQLTGNLTQEARKKHEDVPVIGRKEI
QDAEILARKNPNVLVGDAGVGTAVIEGLAQIIVNGDVPAAIKKKEIVSIDSL
EAGYQSGSFETQNLQIEVKEAGNIILFFDEHQIVGAGATSSDSGSKGLAILKP
ALSREGTLIGATQDQEVNRTILKNAALARNFNEVKVNSAEDTFHLMGRNLYEQ
HHHTLPDNLKAAVDYSIQVIPQRSPLDKAIDLLDMTAAHLAAQHPVTDLTLEI

```

```

AKOKESQKAVAKEDFEKALAAKTRITLQKQIEOHNSONVTATVNDIAESVERLTG
IPYSNMGTDNLERLKGISRLKSHVIGODERAAVAARAIERNRAGFDDGRPTGSFLP
VPGTGKTELAKQLADLFGSKDAIIRLDMSEYNDRATVSKLIGTAGTVGVDNNNN
TLTERVRNRPYAIVLDEIEKADPQIITLLQLVDDGLRDLGGQNTINFRKTVIATS
NAGFGQDQDTSNTESNMDRIAPYFRPEFLNFRNSIIFNHLKQESLESEVLDMLAEVN
QTTAKKGISLITDDAKAHLIDILGYNHANGARPLRIIEQIRDIRITDYVLDHPVKK
LOAILKEGOLVIRQNDQ"
7147. 7830
/gene="rpiA"
/note="Spym3_0608"
7147. 7830
/gene="rpiA"
/note="best blastp hit: gb|AAK33809.1| (AE006538) putative
ribose 5-phosphate isomerase [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative ribose 5-phosphate isomerase"
/protein_id="AAW79215.1"
/db_xref="GI:21904337"
/translation="MEALKKIAGVTAQYVVDGMTIGLTGSTAYYVEEIGRRVKOE
GLQVGVTTSSVTSKQAEVLGILPKSIDDDISLITVDGADVDKKNFGKGGAAALL
MEKIVATPKETIYVVDASKWHEHLGAFKLPVEVQYQADRLFRVFKAGYKPSFRHK
GDSRLVDMONYIIDLGLGCIKDPVAFGHLDDGTGVVVEHGLFNGVMYKIVASKDGV
TVLEAPTAG"
7907. 9118
/gene="deob"
/note="Spym3_0609"
7907. 9118
/gene="deob"
/note="best blastp hit: gb|AAK33810.1| (AE006538) putative
phosphopentomutase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative phosphopentomutase"
/protein_id="AAW79216.1"
/db_xref="GI:21904338"
/translation="MSKFNRIHLVLDVSGIAGRAPADKFEFNAGVADTSDTLCHISE
AAGLSVFNMAKIGLINSRPIPLKTVPTEDNPTGYVTKLEVS LGKRTMTGHWMEIMGL
NITEPEDTFWNGPEEILTKIEEFSGRKIIREANKPYSGTAVIIDDQFPGDMETGELIV
YTSADPVQIQAHEHIIPEVELYKICEYASITLERPALLGRIIARYPDGPGNFTRT
ANRHDYAVSPFOOTVLNKLADAGVPTVAVGKINDIFNGSGITNDMGHKNSHGIDTL
IKTLQLPEFTKGFSTNLVDFDANFGHRRDPESYRDLCHFEFDNRLPIIANMKEDLL

```

```

Query Match 19.6%; Score 114; DB 1; Length 53354;
Best Local Similarity 53.7%; Pred. No. 1.8e-18;
Matches 259; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Qy 1 ATGTTAATGCTTTTATTCATTATTTGGCCTTACTGTAGGCAGCATTTCCATCTGGCTTA 60
    ||| || || || || || || || || || || || || || || || || || || || ||
Db 22122 ATGAATTTACTACTTTTATTTACCATTGCCTTATTTACTTGTGTTCTTATTTCCAACTGGACTA 22063
    ||| || || || || || || || || || || || || || || || || || || || ||

Qy 61 ATTGTGGCAAGCTTGCACAAAGGAATTGATATTCGGGACGACGGAAGCGCAACTTAGGC 120
    ||| || || || || || || || || || || || || || || || || || || || ||
Db 22062 TGGATTGGACAGTACTTTTACCACATCACTTACGAGAGCATGGATCAGGAATACTGGA 22003
    ||| || || || || || || || || || || || || || || || || || || || ||

Qy 121 GCTACCAATGATTCGCTACATTTGGGTGTAAGAGCGTTCGTCGTCATAGCGGAGAT 180
    ||| || || || || || || || || || || || || || || || || || || || ||
Db 22002 ACCACAAATACTTTTCGATTTTAGGTGTCAGGACGAGCAAGCTTACCTAGCTATTGAT 21943
    ||| || || || || || || || || || || || || || || || || || || || ||

Qy 181 ATTTTGAAGGACACATGGCAACTGCATTTGCTTTCTCATGTCATGT---TGATATTCAC 237
    ||| || || || || || || || || || || || || || || || || || || || ||
Db 21942 ATGTTTAAAGGACACATTTCAATATTTGTTACCAATATTTTGGTATGACTTCAATTTCA 21883
    ||| || || || || || || || || || || || || || || || || || || || ||

Qy 238 CCGCTTCTTCGAGGAGTCTTTTCGGCTTTTAGGCCACACGTGTTTCCCATCTTCGCCAAATTT 297
    ||| || || || || || || || || || || || || || || || || || || || ||
Db 21882 TCCATTGCTATCGCTTTTTCGCAGTTTTAGGCATACTTTTCCCTATTTTTCGCCAACTTT 21823
    ||| || || || || || || || || || || || || || || || || || || || ||

Qy 298 AAAGCGGTAAAGCGGTGGCGGACATCAGAGAGCGGTTTGTCTATTTTACGACACCCCTGTTA 357
    ||| || || || || || || || || || || || || || || || || || || || ||
Db 21822 AAAGGTGTAAGGCGGTGACCAAGTCTGCTGTATTGCTAGGCTTTGCTCCCGTTATAT 21763
    ||| || || || || || || || || || || || || || || || || || || || ||

Qy 358 TTTATCAGATGGTTGCGGTATTCTTCATCTTTTATATCTTACTTGACTAAATTTGTTTCTCTC 417
    ||| || || || || || || || || || || || || || || || || || || || ||

```


Db	21762	C T C T T T T T T T A G C A T C T A T C T T T G T T T T A G T T C T C T A T T A T T A T T A G C A T G A T C T T T A	21703
Oy	418	T C A T C G A T G T T A A C A G G A T C T A C T A C T G T A T A T A T A G T T T C T T G C C A T G A T A C G T A T	477
Db	21702	G C T A G T G T G G T T T C A G C T A T C G T G G G T G T A T C T C T T T T A C A T T C C T G C C A T T C A T	21643
Oy	478	T T 479	
Db	21642	T T 21641	

RESULT	14
AF269800/c	
LOCUS	Staphylococcus epidermidis strain SRI clone step.1023g03 genomic sequence.
DEFINITION	3014 bp DNA linear BCT 01-AUG-2000
ACCESSION	AF269800
VERSION	AF269800
KEYWORDS	AF269800.1 GI:9623698
SOURCE	Staphylococcus epidermidis.
ORGANISM	Staphylococcus epidermidis Bacteria; Firmicutes; Bacillales; Staphylococcus. (bases 1 to 3014)
REFERENCE	Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbes,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,p.J.

LOCUS	AX145118	3014 bp	DNA	linear	PAT 31-MAY-2001
DEFINITION	Sequence 3840 from Patent WO0134809.				
ACCESSION	AX145118				
VERSION	AX145118.1	GI:14283683			
KEYWORDS					
SOURCE	synthetic construct.				
ORGANISM	synthetic construct				
REFERENCE	artificial sequences.				
AUTHORS	1 (bases 1 to 3014)				
TITLE	Kimmerly, W.J.				
JOURNAL	Staphylococcus epidermidis nucleic acids and proteins				
FEATURES	Patent: WO 0134809-A 3840 17-MAY-2001;				
source	GLAXO GROUP LIMITED (GB)				
	Location/Qualifiers				
	1. .3014				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="synthetic nucleic acid sequence"				
BASE COUNT	1083 a	444 c	510 g	977 t	
ORIGIN					

Query Match	18.6%;	Score 108.2;	DB 1;	Length 3014;
Best Local Similarity	50.7%;	Pred. No. 5.7e-17;		
Matches 309;	Conservative 0;	Mismatches 273;	Indels 27;	Gaps 1;
Qy 1	ATGTTAAATTCGTTTATTTGGCTACTTGTATAGGACGACATTCCTCTGGCTTA	60		
Db 1223	ATGATGATCATCGTCATGTTAAATCTTTGAGTTATCTGTGTCATTCCTCAAGCGGGTTA	1164		
Qy 61	ATTGTGGGACGCTTCCCAAGGAATTGATATTCGGGACACGGAGCGGCACTTAGGC	120		
Db 1163	ATTATTGGTAAATATTTTTTAAAAAAGATATAAGACAATACGGTAGTGGAAATACTGGA	1104		
Qy 121	GCTACCAATGCATTCCGTACATTTGGGTGTAAAGCTGGTTCGTCGTATAGCCCGAGAT	180		
Db 1103	GCAACTAACAGTTTTTCGTGTTCTTGGAAAGACAGCTGGATTTATAGTACGTTTTTAGAT	1044		
Qy 181	ATTTTGAAGGACACTGGCAACTGCATTTGCCCTTTTCTCATGCATGTTGATATTCACCCG	240		
Db 1043	ATTTTCAAGGATTTATACAGTCTTTTTTCCACTATGTTTCCCGAGTTCATCGCGATGTT	984		
Qy 241	CTTCTT-----GCAGGAGTCTTTTTCGGGTTTTTAGGCCAC	273		
Db 983	GTTATAGCAACCTTCTTTTACAAATGGTTTAATAGTAGGATGTTTTCGCAATACTCGGTGAC	924		

Oy	274	GTGTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCCGTGGCGACATCAGGAGCGTT	333
Db	923	GTGTATCCCAATATATCTGAAATTTAAATGGCGGAAAGCAGTAGCTACCAAGTGCAGGAGTT	864
Oy	334	TTGCTATTTTACGCACCCCTGTTATTTATCAGATGGTTGGGTATTTCTTCATCTTTTA	393
Db	863	GTATTAGGTGTCATCCTATTTTACTTCTTATCTTGGCAATTATCTTTTAGTGTATTA	804
Oy	394	TACTTGACTAAATTTGTTCTCTCTCATCGATGTTAACAGGATCTATACGTGTATATAT	453
Db	803	AAAATCTTTAAATATGTTCTTTTATCAAGTATCATTCAGCAATTAGTTGTGATTGGT	744
Oy	454	AGTTCTTTGTCCATGATACGTATTTATTTGATTCGTTACCCCTGCTCACTATTTTGTG	513
Db	743	TCAATCATCATCTCATGATATATTTTACTTCTGTAGCGAATTTGTTCAATCATATTA	684
Oy	514	ATATACAGACACCGAGCGCAACATTAAACGAATTATCAATAAAACAGAACCTAAAGTAAAA	573
Db	683	ATAATTCGACACAATCTAATATAGTTAGAAATTTTAAAGGAGAAGAACCTAAATTA	624
Oy	574	TGGTTATAA	582
Db	623	TGGATGTA	615

Search completed: December 25, 2002, 10:52:49
 Job time : 10757.6 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 136.703 Seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-4
Perfect score: 582
Sequence: 1 atqtaattgccttattaat.....ctaaqtaaatggttataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

```
Minimum DB seq length: 0.
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	582	100.0	582	20	AAZ20371	B. subtilis B-vnes
2	582	100.0	582	21	AAA09182	B. subtilis B-vnes
3	176	30.2	495269	24	ABQ67195	Listeria innocua c
4	117.4	20.2	636	24	ABNS7419	Listeria innocua c
5	115.8	18.9	621	24	ABN70980	Streptococcus poly
6	114	19.6	720	24	ABN67450	Streptococcus poly
c 7	109.6	18.8	6691	20	AAH13025	Enterococcus faeca
8	108.2	18.6	615	24	ABNS2939	Staphylococcus epi
c 9	108.2	18.6	3014	22	AAHS4476	S. epidermidis gen

C	10	108.2	18.6	4105	22	AAH54115	S. epidermidis gen
	11	106.6	18.3	594	22	AAH53731	S. epidermidis ope
	12	101.6	17.5	642	23	AA555594	Streptococcus pne
	13	101.6	17.5	642	23	AA555801	Streptococcus pne
	14	100	17.2	642	21	AAAO9181	S. pneumoniae S-yn
	15	100	17.2	6171	19	AA522170	Streptococcus pne
	16	90.6	15.6	609	23	AA554402	Streptococcus pne
	17	88.2	15.2	606	23	AA551868	Staphylococcus aur
	18	80.4	13.8	818	18	AAV75059	Staphylococcus aur
C	19	76.6	13.2	344	20	AAZ20352	Gene encoding bact
C	20	55.2	9.5	1657	24	ABQ70507	Listeria monocytog
	21	48	8.2	600	23	AA553263	Haemophilus influe
	22	48	8.2	1830121	17	AA742063	Haemophilus influe
	23	46.4	8.0	72	24	ABK76809	Bacillus lichenifo
C	24	45.2	7.8	556	23	ABV40063	Human prostate exp
C	25	45.2	7.8	556	23	ABV40163	Human prostate exp
C	26	45.2	7.8	556	23	ABV42105	Human prostate exp
C	27	45.2	7.8	556	23	ABV43601	Human prostate exp
	28	43.4	7.5	648	24	ABQ48108	Oligonucleotide fo
C	29	43.4	7.5	648	24	ABQ48109	Oligonucleotide fo
	30	43	7.4	789	23	AA553719	Helicobacter pylor
	31	43	7.4	823	19	AAAX1476	H. pylori GHPO 108
C	32	42.6	7.3	309	23	ABV44994	Human prostate exp
C	33	42.4	7.3	612	23	AA556022	Human prostate exp
C	34	42.2	7.3	10205	24	ABL70236	Salmonella typhi D
C	35	42.2	7.3	10205	24	ABK31875	Chemically treated
	36	42	7.2	9905	24	ABL32062	Signal transductio
	37	41.8	7.2	6956	24	ABL70225	Human immune syste
	38	40.8	7.0	696	22	AAAL14782	Chemically treated
	39	40.2	6.9	626	24	ABQ56717	Human breast cancer
C	40	40.2	6.9	752	24	ABQ14854	Human colon cancer
	41	40.2	6.9	752	24	ABQ14855	Oligonucleotide fo
C	42	39.8	6.8	663	23	AA553880	Oligonucleotide fo
C	43	39.6	6.8	394	22	AAAL10020	Helicobacter pylor
C	44	39.6	6.8	476	23	ABV58368	Human breast cancer
C	45	39.6	6.8	6121	24	ABL92201	Human prostate exp

ALIGNMENTS

RESULT 1

AAZ20371

ID AAZ20371 standard; DNA; 582 BP.

XX

AC AAZ20371;

XX

DT 17-NOV-1999 (first entry)

XX

DE B. subtilis B-ynes protein coding sequence

XX

KW General essential protein; pathogenic bacter

KW bacterial growth; B-ynes; ds.

XX

OS *Bacillus subtilis*.

XX
XX

FH	Key	Location/Qualifiers
100	100	100
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129	129	129
130	130	130
131	131	131
132	132	132
133	133	133
134	134	134
135	135	135
136	136	136
137	137	137
138	138	138
139	139	139
140	140	140
141	141	141
142	142	142
143	143	143
144	144	144
145	145	145
146	146	146
147	147	147
148	148	148
149	149	149
150	150	150
151	151	151
152	152	152
153	153	153
154	154	154
155	155	155
156	156	156
157	157	157
158	158	158
159	159	159
160	160	160
161	161	161
162	162	162
163	163	163
164	164	164
165	165	165
166	166	166
167	167	167
168	168	168
169	169	169
170	170	170
171	171	171
172	172	172
173	173	173
174	174	174
175	175	175
176	176	176
177	177	177
178	178	178
179	179	179
180	180	180
181	181	181
182	182	182
183	183	183
184	184	184
185	185	185
186	186	186
187	187	187
188	188	188
189	189	189
190	190	190
191	191	191
192	192	192
193	193	193
194	194	194
195	195	195
196	196	196
197	197	197
198	198	198
199	199	199

ET	CDS	ET	CDS
1.1582	1.1582	1.1582	1.1582

```

FF      /* tag= a
FE      /a/

```

```

F#F
yy
/product= B-yne$

```

XX
PN W00033871-22

PN W099338/1-A2.
YY

08-JUL-1999
XX
PD

FD 08-JUL-1999.
XX

30-DEC-1998: 98WQ-US27918.

FF 30 DEC 1990; 30MC-US27310.
XX

DR WPI: 1999-430230/36.
DR P-PSDB: AAY22580.
XX Streptococcus pneumoniae general essential protein genes and proteins,
PT useful for identification of antibacterial agents.
XX Disclosure; Fig 24; 124pp; English.
XX This sequence encodes the Bacillus subtilis B-ynes protein. B-ynes is
CC related to the Streptococcus pneumoniae general essential
CC protein (GEP) gene of the invention. The genes encoding the GEP
CC polypeptides are useful molecular tools for identifying similar genes in
CC pathogenic microorganisms, such as pathogenic strains of Bacillus. In
CC addition, the operons containing genes encoding GEP and the polypeptides
CC themselves, are useful targets for identifying compounds that are
CC inhibitors of the pathogens in which the GEP are expressed. Such
CC inhibitors are useful for inhibiting bacterial growth by being
CC bacteriostatic or bacteriocidal.
XX
XX Sequence 582 BP; 147 A; 115 C; 122 G; 198 T; 0 other;
XX
Query Match 100.0%; Score 582; DB 20; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.6e-158;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTTAATGCTTTATTCATATTTTGGCCTACTTGATAGGCACCATTCCTCTGGCTTA 60
Db 1 ATGTTAATGCTTTATTCATATTTTGGCCTACTTGATAGGCACCATTCCTCTGGCTTA 60
Qy 61 ATTGGGCAAGCTTGCCAAAGGAATTCATATTCGGGAGACGGAAGCGCACTTAGGC 120
Db 61 ATTGGGCAAGCTTGCCAAAGGAATTCATATTCGGGAGACGGAAGCGCACTTAGGC 120
Qy 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Db 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Qy 181 ATTTTGAAGGACACATGCGCACTGCAATTCGCTTTCTCATGCAATGATATTCACCCG 240
Db 181 ATTTTGAAGGACACATGCGCACTGCAATTCGCTTTCTCATGCAATGATATTCACCCG 240
Qy 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Db 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Qy 301 GCGGTAAGCCGTCGGGACATCAGAGGCGTTTGGTATTTTACGACCCCTGTTATTT 360
Db 301 GCGGTAAGCCGTCGGGACATCAGAGGCGTTTGGTATTTTACGACCCCTGTTATTT 360
Qy 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATGTTTCTCTCTCTCA 420
Db 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATGTTTCTCTCTCTCA 420
Qy 421 TCGATGTTTAAACAGGATCTATACCTGTTATATATAGTTTCTTGTCCATGATACGTTTA 480
Db 421 TCGATGTTTAAACAGGATCTATACCTGTTATATATAGTTTCTTGTCCATGATACGTTTA 480
Qy 481 TTGATGCTGTTACCTGCTCAGTATTTTGTGATATACAGACACCGAGCAATTTAAA 540
Db 481 TTGATGCTGTTACCTGCTCAGTATTTTGTGATATACAGACACCGAGCAATTTAAA 540
Qy 541 CGAATTATCAATAAACAAGACCTTAAAGTAAATGGTTATATA 582
Db 541 CGAATTATCAATAAACAAGACCTTAAAGTAAATGGTTATATA 582
RESULT 2
AAA09182
ID AAA09182 standard; DNA: 582 BP.
XX
AC AAA09182;
XX
DT 10-AUG-2000 (first entry)

XX B. subtilis B-ynes coding sequence.
DE
XX B-ynes; S-ynes; survival; antibacterial; inhibitor; ds..
XX Bacillus subtilis.
OS
XX WO200020527-A1.
PN
XX 13-APR-2000.
PD
XX 30-SEP-1999; 99WO-US22665.
PF
XX 30-SEP-1998; 98US-0163445.
PR
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX Fritz C, Youngman P, Guzman L;
PI
XX WPI: 2000-303799/26.
DR
XX P-PSDB: AAY92247.
XX
Methods for identifying an antibacterial agent for treating
PT Streptococcus pneumoniae infections comprises detecting an interaction
PT between a ynes polypeptide and a test compound
XX
XX Disclosure; Fig 2; 65pp; English.
XX
XX Bacillus subtilis B-ynes is a homologue of the Streptococcus pneumoniae
CC ynes gene (S-ynes) which is essential for survival for a wide range of
CC bacteria. Identifying an antibacterial agent comprises contacting a ynes
CC polypeptide (S-ynes) with a test compound and detecting an interaction of
CC the test compound with the S-ynes polypeptide which indicates that the
CC compound is an antibacterial agent. Alternatively, detecting a decrease
CC in function of the polypeptide contacted with the test compound and
CC determining whether the compound inhibits growth of bacteria, relative to
CC the growth of bacteria cultured in the absence of a test compound where
CC inhibition of growth indicates the compound is an antibacterial agent.
CC Inhibitors of S-ynes function are useful for treating a S. pneumoniae
XX infection in mammals.
XX
XX Sequence 582 BP; 147 A; 115 C; 122 G; 198 T; 0 other;
XX
Query Match 100.0%; Score 582; DB 21; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.6e-158;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTTAATGCTTTATTCATATTTTGGCCTACTTGATAGGCACCATTCCTCTGGCTTA 60
Db 1 ATGTTAATGCTTTATTCATATTTTGGCCTACTTGATAGGCACCATTCCTCTGGCTTA 60
Qy 61 ATTGGGCAAGCTTGCCAAAGGAATTCATATTCGGGAGACGGAAGCGCACTTAGGC 120
Db 61 ATTGGGCAAGCTTGCCAAAGGAATTCATATTCGGGAGACGGAAGCGCACTTAGGC 120
Qy 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Db 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Qy 181 ATTTTGAAGGACACATGCGCACTGCAATTCGCTTTCTCATGCAATGATATTCACCCG 240
Db 181 ATTTTGAAGGACACATGCGCACTGCAATTCGCTTTCTCATGCAATGATATTCACCCG 240
Qy 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Db 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Qy 301 GCGGTAAGCCGTCGGGACATCAGAGGCGTTTGGTATTTTACGACCCCTGTTATTT 360
Db 301 GCGGTAAGCCGTCGGGACATCAGAGGCGTTTGGTATTTTACGACCCCTGTTATTT 360
Qy 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATGTTTCTCTCTCTCA 420
Db 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATGTTTCTCTCTCTCA 420

```
Db 361 ATCAGATGTTGGGTAATCTTCACTCTTTTATATCTGACTAAATTTGTTCTCTCTCA 420
Qy 421 TCGATGTTAAACAGGATCTACTGTTATATATAGTTCTTTGTCCTCATGATAGTATTTA 480
Db 421 TCGATGTTAAACAGGATCTACTGTTATATATAGTTCTTTGTCCTCATGATAGTATTTA 480
Qy 481 TTGATGTCGTTACCCCTGCTCACTATTTTGTGATATACAGACACCGAGCAACATTAA 540
Db 481 TTGATGTCGTTACCCCTGCTCACTATTTTGTGATATACAGACACCGAGCAACATTAA 540
Qy 541 CGAATTAATCAATAAACAACACCTAAAGTAAATGGTTATA 582
Db 541 CGAATTAATCAATAAACAACACCTAAAGTAAATGGTTATA 582

RESULT 3
ABQ67195
ID ABQ67195 standard; DNA: 495269 BP.
XX
AC ABQ67195;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #8.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI: 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
XX
PS Claim 5; SEQ ID 8; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ671212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 495269 BP; 159529 A; 88849 C; 93135 G; 153754 T; 2 other;
Query Match 30.2%; Score 176; DB 24; Length 495269;
Best Local Similarity 57.9%; Pred. No. 8.8e-40;
Matches 336; Conservative 0; Mismatches 235; Indels 9; Gaps 1;

Qy 12 TTTATGATTAATTTGGCTTACTTGATAGGACGATTCATCTGGCTTAATTTGGGCAA 71
Db 441809 TTTGCTTTCTTGTAGCTTATGTAATCGGTTCAATACCTTCTGCTTATGGATCGGTA 441868
```

```
Qy 72 GCTGCCAAAGGAATGATATCGGGACGACGGAAGCGCAACTTAGCGCTACCAATGC 131
Db 441869 AATTTTTTACAAAGATATTCGTGATTTTGGTAGCGGAATTTAGGAGCTACAACTC 441928
Qy 132 ATTCCGTACATTTGGGTGCTGCTGCGTGCATAGCGGGAGATATTTTGAAGG 191
Db 441929 CTTCCGCTGTTTGGGGTAAAGCGGCAGTATTTGTACAGTGATGATATTTTAAAGG 441988
Qy 192 GACACTGGCAACTGCCATTGCTTTTCTCATGCTGTTGATATCA-----CCCGCT 242
Db 441989 TACTGTGCGCAACATTACTTCCATTTCTTTTCCAACTAAATGTTAATCATATTTCTGGTT 442048
Qy 243 TCTTGACGAGAGTCTTTTGGCGTTTATAGGCCACGCTTTTCCCATCTTCCGCAAAATTTAAAGG 302
Db 442049 ACTTACGGGAGCATTTGCGATTAATCGACATAGTTTCCCGCTTTTCCGAGGATTCAGAGG 442108
Qy 303 CGGTAAAGCCGTGGCGACATCAGAGAGCGGTTTGTCTATTTTACGACACCCCTGTTATTTAT 362
Db 442109 TGGGAAAGCAGTTGCTACTTCTGCTGCTGTTATCTTCTTATTAACATTTAAATAATCAGCAAAATATGTCGCTTAGTTC 442168
Qy 363 CACGATGTTGGCGTATTTCTTCATCTTTTATCTACTGACTAAATTTGTTCTCTCTCATC 422
Db 442169 AGCAGCACTAGTCTGTTTCTTATTAACATTTAAATAATCAGCAAAATATGTCGCTTAGTTC 442228
Qy 423 GATGTTAAACAGGATCTATATCTGTTATATAGTTTCTTTTGTCCATCATGATATTTTATT 482
Db 442229 AATGTTGGAGCCTTCCACCATTAATTAATTTCTTTTTCATGGGAGACTGGATTTTAAAT 442288
Qy 483 GATTGCTGTACCCTGCTCCTACTATTTTGTGATATACAGACACCGAGCAACATTTAAAGG 542
Db 442289 TATCCTCGTTGCTTGCATCTCTCTTTTGTATTGTCGCCACCGAGCAAAATATTACGCG 442348
Qy 543 AATTATCAATAAACAACAGCACTTAAAGTAAATGGTTATA 582
Db 442349 CATTCGAACGCGGAGAACAAACAAATAAATAATGGATGTAA 442388
```

RESULT 4

```
ABN67449
ID ABN67449 standard; DNA: 636 BP.
XX
AC ABN67449;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 2811.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
WPI: 2002-352536/38.
DR P-PSDB; A3P26818.
XX
```


Db 122 GTATTTAGGTGTAAGAGTGGCAATGTCATCTTTGACTATTGACATTTTAAAGGGGACTC 181
Qy 197 TGGCAACTGCAATGCCCTTCTCATGATGTTGATA---TTCACCCCGCTCTTCGAGGAG 253
Db 182 TAGCAACACTTATCCCTATCATATAGGTATACGACAGATCTCCATTTTATCGGTT 241
Qy 254 TCTTTGGGTTTATAGGCCAGGTGTTCCCATCTTCGCCAAATTTAAAGGCGGTAAGCCG 313
Db 242 TTTTGTCTATATAGGTACACACTTTCCTATTTTGCACAAATTCAAAGGTGGGAAGCCG 301
Qy 314 TGGCGACATCAGGAGGCGTTTGTCTATTTTACGACACCCCTGTTATTTATCAGATGTTG 373
Db 302 TTGCTACAAGCGCTGGGGTTTGTAGGATTTGCACCTTCTTTTCTTACTTATTTAG 361
Qy 374 CGGTATCTTCATCTTTTATACCTTATCTTCTCTCTCATCGATTTTACAG 433
Db 362 TTATCTTTTATTAACGCTTATCTTTTACGATGATTTCCCTATCCAGTATTTACAGTAG 421
Qy 434 -GGATCTATACCTGTTATATATAGTTTCTTTGTCATGATAGGTATTTAT-----TGATG 487
Db 422 CTGTTGTAGGTATCTTAGTGTCTTAATTTTCCATTAGTAGGCTTTATATATTAACGGACT 481
Qy 488 TCGTTACCTGCTCAGTATTTTGTGATATACAGACACCGGACGCAACATTTAAAGCAATTA 547
Db 482 ACGACTGGATATTTACCACTGTGTTATCTTAATGCGCCTAACGATTTATTCGAGATC 541
Qy 548 TCAATAAAGACAGAACCTTAAAGTAAATGTTTATTA 582
Db 542 AGGATAATATCAACAGTATTCGAAAGAGCGAAGAA 576

RESULT 6

ABN67450
ID ABN67450 standard; DNA; 720 BP.
XX
AC ABN67450;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 2813.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS Streptococcus pyogenes.
XX

WO200234771-A2.

XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI: 2002-352536/38.
XX P-PSDB: ABP26819.
XX

XX New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein .

PS Claim 7; Page 3432; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

SQ Sequence 720 BP; 199 A; 123 C; 112 G; 286 T; 0 other;

Query Match 19.6%; Score 114; DB 24; Length 720;
Best Local Similarity 53.7%; Pred. No. 6e-23;
Matches 259; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Qy 1 ATCTTAATGCTTTTATTTGCTTATTTTGGCTTACTTTGATAGGCAGCATTTCCATCTGGCTTA 60
Db 82 ATGAATTTACTACTTTTATTTATTTTACCATTGCTTATTTTACTTGGTTCTATTCCAACTGGACTA 141
Qy 61 ATTGTGGGCAAGCTTGCCTTATTTGCTTATTTTGGCTTACTTTGATAGGCAGCATTTCCATCTGGCTTA 120
Db 142 TGGATTGGCAGAGTACTTTTACCACATCACTTACGAGAGCATGGATCAGGAATACTGGA 201
Qy 121 GCTACCAATGCTTCCGTACATTTGGGTGTAAAGCTGGTTCGGTGCATAGCCGGAGAT 180
Db 202 ACCACAAATACTTTTCGGATTTTAGGTGTCAAGCAGGACAGCTACTTACGTATTGAT 261
Qy 181 ATTTTGAAGGACACTGGCAACTGCATTTGCTTTTCTCATGCTGT---TGATATTTCAC 237
Db 262 ATGTTTAAAGGGACACTTTTCAATATTCTTACCAATTATTTTGTGTATGACTTCAATTCA 321
Qy 238 CCGTCTCTTCGAGGAGTCTTTTGGGTTTATAGGCCACGTTTCCCATCTTCGCCAAATTT 297
Db 322 TCCATTGCTATCGGCTTTTTCGAGTTTATAGGGCATACTTTTCTTATTTTGGCAACTTT 381
Qy 298 AAAGCGGTAAAGCGTGGCGACATCAGAGGCGCTTTTGTCTATTTTACGACACCTCTGTA 357
Db 382 AAAGTGTGTAAAGCGGTAGCAACAAAGTGTGGTGTATTGCTAGGCTTTGCTCCGTTATAT 441
Qy 358 TTTATCAGATGTTGCGGTATTTCTTCTATCTTTTATATCTTGACTAAATTTGTTCTCTC 417
Db 442 CTCCTTTTATAGCATCTATCTTTTGTGTTTGTAGTCTCTCTATTTATTTAGTATGATATCTTTA 501
Qy 418 TCATCGATGTTTACAGGATCTATATCTGTTATATATATAGTTTCTTTGTCATGATACGTAT 477
Db 502 GCTAGTGTGTTTTCAGCTATCTGTTGGTGTGTATCTGTTTATCTGTTTAAACATTTTCTCGCATCAT 561
Qy 478 TT 479
Db 562 TT 563

RESULT 7

AAI13025/C
ID AAI13025 standard; DNA; 6691 BP.

XX
XX AAI13025;
XX AC AAI13025;

DT 19-MAR-1999 (first entry)

XX Enterococcus faecalis genome contig SEQ ID NO:88.
XX

XX Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-0508985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

XX Claim 1; Page 592-595; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.

XX SQ Sequence 6691 BP; 2383 A; 1085 C; 1355 G; 1858 T; 10 other;

Query Match 18.8%; Score 109.6; DB 20; Length 6691;
Best Local Similarity 53.4%; Pred. No. 2.7e-21;
Matches 254; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

QY 1 ATGTTAATTCCTTATTGATATTGTCCTACTTGTATAGGACGATTCATCTGGCTTA 60

Db 5424 ATGAAATCGTATTTTGTACTTGTGCTATTTATTTAGGTTTCGATTCCTCAGGTGTT 5365

QY 61 ATTGTGGGCAAGCTTCCAAAGGANTGTATTTGGGAGCAGGAGCGCACTTAGGC 120

Db 5364 TGGATTGGTAAACTTTTCTTTTAAAGAGATATACGCCAATTTGGGAGTGGGAATACAGGA 5305

QY 121 GCTACCAATCATTCGCTACATTTGGGTGTAAGCTGGTTCGTCATAGCGGAGAT 180

Db 5304 AACACCAATACATTCGTGCTTAGGGAACCTGCCGGAATTTAGGTATTTAATTCGAT 5245

QY 181 ATTTTGAAGGACACTGGCAACTGCATTCGCTTTTCTCATGCATGT---TGATATTAC 237.

Db 5244 ATCTTGAAGGACGTTAGCCACTTCATTACCCCTATTGTTTGGTTTACAGGCGGTGAAT 5185

QY 238 CCGTTCCTTCAGAGAGTCCTTGGGGTTTATAGGACAGTGTTTCCCATCTTCGCGCAATTT 297

Db 5184 CCGCTCTTCTTGGGTAGCAGCTGTTTATAGGCATACCTTCCCTATTTTGGCAAATTC 5125

QY 298 AAGGCGGTAAAGCCGTGGCGACATCAGGAGCGTTTGTGATATTTACGACCCCTGTTA 357

Db 5124 AAGGCGGTAAAGCCGTGGCGACATCAGGCGTGGCATGTTTATGATACAGACCCCATTT 5065

QY 358 TTTATCAGATGGTTCGGGTATTCTTCATCTTTTATACCTTACTAAATTTGTTTCTCTC 417

Db 5064 TTTATTTATCTGCTCTTATTTTGTGATTTGTCTGTATCTACCATGATGGTGAAGTTA 5005

QY 418 TCATCGATGTTAACAGGGATCTATCTGTATATATAGTTCTTTGTCATGATAC 473
Db 5004 ACAAGTATGATTAGCGCTCTACTAATTACACTTCTTACTATTATTTGGCTTTAC 4949

RESULT 8

ABN92939

ID ABN92939 standard; DNA; 615 BP.

XX AC ABN92939;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2402.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

XX 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX P-PSDB; ABP40394.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -

XX Disclosure; SEQ ID 2402; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (OFF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP3960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.XX N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.

XX SQ Sequence 615 BP; 185 A; 81 C; 112 G; 237 T; 0 other;

Query Match 18.6%; Score 108.2; DB 24; Length 615;
Best Local Similarity 50.7%; Pred. No. 2.7e-21;
Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

QY 1 ATGTTAATTCCTTATTGATATTGTCCTACTTGTATAGGACGATTCATCTGGCTTA 60

Db 7 ATGATGATCATCGTCATGTTAATCTTTGAGTTATCTGATGTTCCCAAGCGGTTA 66

QY 61 ATTGTGGCAAGCTTCCCAAGGAATTGATATTTCGGGAGCAGGAGCGCAACTTAGGC 120

Db 67 ATATTGGTAAATTTATTTTAAAGAGATATAAGCAATACGGTAGTGGAAATCTGGA 126

QY 121 GTACCAATTCATTCGTTACATTTGGGTGTAAAGCTGGTTCGTCGTATACCCGAGAT 180

Db 127 GCAACTAACAGTCTTCGTTCTTGGAGACAGCTGGATTATAGTTACGTTTTTATAGAT 186

QY 181 ATTTTGAAGGACACTGCAACTGCAATTCCTTCTCATGCATGTTGATTTACCCG 240

Db 187 ATTTTCAAGGATTTATTACAGTCTTTTCCCACTATGTTCCAGGTTTCATCGGATGGT 246
 Qy 241 CTTCCTT-----GCAGGACTCTTTCGGGTTTTCAGGCCAC 273
 Db 247 GTTATAAGCACCTCTTTTACAAATGGTTTAAATAGTAGGATGTTTGCATCTCGGTTCAC 306
 Qy 274 GTGTTTCCCATCTTCGGCCAAATTTAAAGCGGTAAGCCGTGGCGACATCAGGAGGCGTT 333
 Db 307 GTGTATCCATATATCTGAATTTAATGGCGGAAGACAGTACCTACAGTGCAGGAGTT 366
 Qy 334 TTGCTATTTTACGACCCCTGTTTATTTATCAGGATGTTTCGGGTTTTCATCTTTTAA 393
 Db 367 GTATTAGGTGCAATCTTATTTACTTCTTATCTTGGCAATTTATCTTTTTCAGTGTATTA 426
 Qy 394 TACTTGACTAAATTTGTTCTCTCATCATGTTTACAGGGATCTATCTATATATAT 453
 Db 427 AATATCTTTAAATATGTTTCTTTATCAAGTATCATTCGACGAATTTAGTGTGATGGT 486
 Qy 454 AGTTCTTTTGCATGATAGTATTTTATGATGTTCTGTTACCCCTGCTCACTATTTTGTG 513
 Db 487 TCATCATCATTCATGATTTATTTTACTTGTGTTAGCGGAATTTGTTCAATCATATTA 546
 Qy 514 ATATACAGACACCGAGCGAATTAAGCAATTAATCAATTAACAGAACCTAAAGTAAA 573
 Db 547 ATAATTCGACACAAATCTAATATAGTTAGAATTTTAAAGGAGAAGAACCTAAAGTAAA 606
 Qy 574 TGGTTATATA 582
 Db 607 TGGATGTAA 615

RESULT 9
 AAH54476/C
 ID AAH54476 standard; DNA: 3014 BP.
 XX
 AC AAH54476;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3840.
 XX
 KW Staphylococcus epidermidis S1 strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 XX
 XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 8; Page 1482-1483; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 3014 BP; 1083 A; 444 C; 510 G; 977 T; 0 other;

Query Match 18.6%; Score 108.2; DB 22; Length 3014;
 Best Local Similarity 50.7%; Pred. No. 5.1e-21;
 Matches 3C9; Conservative 0; Mismatches 273; Indels 27; Gaps 1;
 Qy 1 ATGCTAATTCGCTTTATTGATTATTTTGGCTACTTGTAGGACAGCATTCATCTGCTTA 60
 Db 1223 ATGATGATCATCGTCATGTTAATCTTGGAGTTATCTGATTTGGTGCATTCACAGCGGTTA 1164
 Qy 61 ATTGTGGGCAAGCTTCCCAAGGAATTTGATTTGGGAGACACGGAAGCGCAACTTAGGC 120
 Db 1163 ATTATTGGTAAATATTTTAAAAAAGATATAAGACAATACGCTAGTGGAAATACTGGA 1104
 Qy 121 GCTACCAATGCATTCGTCATGTTGGGTGTAAGAGCTGGTTCGCTGCATAGCGGAGAT 180
 Db 1103 GCAACTAACAGTCTTTCGTTCTTGGAAAGACAGCTGGATTTATAGTACGTTTATAGAT 1044
 Qy 181 ATTTGAAGGACACTGCGCACTGCATTCCTTTCTCATGCTGTTGATATTCACCCG 240
 Db 1043 ATTTCAAGGATTTATACAGTCTTTTTCACACTATGTTCCAGCTTCATCGGATGGT 984
 Qy 241 CTTCCTT-----GCAGGACTCTTTCGGGTTTTCAGGCCAC 273
 Db 983 GTTATAAGCACCTTCTTTTACAAATGGTTTAAATAGTAGGATTTTGCATCTCGGTTCAC 924
 Qy 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAGCCGTGGCGACATCAGGAGGCGTT 333
 Db 923 GTGTATCCAAATATATCTGAATTTTAAATGGCGGAAGACAGTACGTACCTACAGTGCAGGAGTT 864
 Qy 334 TTGCTATTTTACGACCCCTGTTTATTTATCAGCATGTTTCGGGTTATTTCTTCATCTTTTAA 393
 Db 863 GTATTAGTGTCAATCTTATTTACTTCTTATCTTGGCAATTTATCTTTTTCAGTGTATTA 804
 Qy 394 TACTTGACTAAATTTGTTCTCTCATCGATGTTAACAGGATCTATATCTGTATATATAT 453
 Db 803 AAATCTTTAAATATGTTTCTTTTATCAAGTATCATTTGCAGCAATTTAGTTCGTGATTTGT 744
 Qy 454 AGTTTCTTTTGCATGATGATGTTTATTTATGATTTGCTGCTTACCCTGCTCACTATTTTGTG 513
 Db 743 TCAATCATCATTCATGATTTATTTTACTTGTCTAGCGGAATTTGTTCAATCATATTA 684
 Qy 514 ATATACAGACACCGAGCGAATTTAAAGCAATTTATCAATTAACAGAACCTAAAGTAAA 573
 Db 583 ATATTCGACACAAATCTAATATAGTTAGAATTTTAAAGGAGAAGAACCTAAAGTAAA 624
 Qy 574 TGGTTATATA 582
 Db 623 TGGATGTAA 615

RESULT 10
 AAH54115/C
 ID AAH54115 standard; DNA: 4105 BP.
 XX
 AC AAH54115;
 XX
 DT 03-SEP-2001 (first entry)
 XX

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3479.
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 XX
 OS Staphylococcus epidermidis.
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis .
 XX
 PS Claim 8; Page 1047-1049; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AG81454 to AG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to
 CC AAH5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 4105 BP; 1478 A; 590 C; 744 G; 1293 T; 0 other;
 Query Match 18.6%; Score 108.2; DB 22; Length 4105;
 Best Local Similarity 50.7%; Pred. No. 5.7e-21;
 Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;
 QY 1 ATGTTAATGCTTATTGATTATTTTGGCCCTAGTTAGGACGATTCATCTGGCTTA 60
 DB 1557 ATGATGATCATCGTCATGCTTAATCTTGAGTTATCTGATGTGCTATCCCAAGCGGGTTA 1498
 QY 61 ATTGGGCGACCTGCCAAGGAATGATATCGGGAGCAGCGAAGCGCACTTAGGC 120
 DB 1497 ATTATTGTAATATTTTAAAAAAGATATAAGACAATACGGTAGTGGAAATACTGGA 1438
 QY 121 GCTACCAATGATCCGTFACATTTGGGTGTAAGAGCTGGTTCGGTCGTCATAGCGGGAGAT 180
 DB 1437 GCACCTACAGTTTTCGGTGTCTTGGGAAGACCGACTGGATTTATGATACGTTTTTAGAT 1378
 QY 181 ATTTTGAAGGACACTGCTGCACTGCTTTCTCATGCACTGTGATATTCACCCG 240
 DB 1377 ATTTTCAAGGATTTATACAGTCTTTTCCACTATGTTCCAGTTCATGCGGATGGT 1318
 QY 241 CTCTCTT-----CGAGGAGTCTTTGGGGTTTTAGGCCAC 273
 DB 1317 GTTATAAGCACCTTCTTTACAAATGGTTTAAATAGTAGGATTTGTTGCAATACTCGGTCAC 1258

QY 274 GTCTTTCCCATCTTCGCCAAATTTAAGCGCGTAAAGCCGTCGGACATCAGAGCGGTT 333
 DB 1257 GTGTATCCCAATATATCTGAATTTAATGGCGAAAGACAGTACTACCACTGCAGGAGTT 1198
 QY 334 TTGCTATTTTACGACCCCTGTTTATTTATCAGCATGGTTCGGGTATTTCTTCATCTTTTAA 393
 DB 1197 GTATTAGGTGTCATCTTATTTTACTTCTTATCTTGGCAATATCTTTTTTAGTGATTA 1138
 QY 394 TACTTGACTAAATTTGTTCTCTCTCATCGATGTTACAGGAGGATCTATACGTATATAT 453
 DB 1137 AAAATCTTTAAATATGTTTCTTTATCAAGTATCATTTGACGCAATATTAGTGTGATGGT 1078
 QY 454 AGTTCTTGTCCATGATACGTATTTATTTATGATTGCTTACCCTGCTCAGTATTTTGTG 513
 DB 1077 TCAATCATCATTCATGATATATTTTACTTGTCTGAGCGGAATTTGTTCAATCATATTA 1018
 QY 514 ATATACAGACACCGAGCGAATTAACAGAAATTAATCAATAAAGACAGACCTAAAGTAAA 573
 DB 1017 ATRATTCGACACAAATCTAATATATAGTTAGAAATTTTAAAGGAGAGAAGCACTAAATAAA 958
 QY 574 TGGTTATAA 582
 DB 957 TGCATGTAA 949
 RESULT 11
 AAH53731
 ID AAH53731 standard; DNA; 594 BP.
 XX
 AC AAH53731;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2855.
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 KW
 XX Staphylococcus epidermidis.
 XX WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 DB P-PSDB; AG82881.
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 XX useful for vaccinating against infections, e.g. endocarditis .
 PS Claim 8; Page 744; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AG81454 to AG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to

CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 594 BP; 181 A; 78 C; 108 G; 227 T; 0 other;

Query Match 18.3%; Score 106.6; DB 22; Length 594;
 Best Local Similarity 50.9%; Pred. No. 7.7e-21;
 Matches 302; Conservative 0; Mismatches 264; Indels 27; Gaps 1;

QY 17 TGATTATTTTGGCCCTACTTGTAGGACGATTCCTCATCTGGCTTAATTTGGCGAAGCTTG 76
 DB 2 TGTAACTCTGATGATCTGATGTCATCCCAAGCGGTAAATTTGGTAAATAT 61
 QY 77 CCAAGGAATGATATTCGGGAGCAGGAAAGCGCACTTAGCGCTACCAATGCAATCC 136
 DB 62 TTTTAAAGAGATATAAGACAATACGCTAGTGGAAATACCTGGAGCACTAACAGTTTC 121
 QY 137 GTACATTTGGGTAAAGCTGTTCCGTCGTCATAGCCGGAGATATTTGAAAGGACAC 196
 DB 122 GTGTTCTTTGGAAGACCACTGGATTTATAGTTACCTTTTATAGATATTTTCAAGGATTTA 181
 QY 197 TGGCAACTGCATTCCTTCTCATGTCATGTTGATATTACCCGCTTCTT----- 246
 DB 182 TTACAGTCTTTTCCACATAGTGGTCCAGATTCATCGGAGGTGTTATAGCACTTCT 241
 QY 247 -----CGAGGACGTCCTTTGCGGTTTATAGGACGCTGTTCCCATCTTCG 289
 DB 242 TTACAAATGGTTTAAATAGTAGGATGTTTGGCAATCTCGTCACGCTATCCAAATATC 301
 QY 290 CCAATTTAAAGCGGTAAAGCGTGGCAGATCAGGAGCGGTTTGGCTATTATAGGTC 349
 DB 302 TGAATTTTAAAGCGGAAAGCAGTAGTACCAGTGCAGAGGTGTTATAGGTGTCATC 361
 QY 350 CCTGTTTATTTATCAGATGTTGGGTTATCTTCTCATCTTTTATCTTATGCTAAATTTG 409
 DB 362 CTATTTTACTTCTTCTTGGCAATATCTTTTATGTCATTTAAATCTTTTAAATATG 421
 QY 410 TTTCTCTCTCATGATGTTAAAGGAGTCTATCTGTTATATATAGTTCTTTTGTCCATG 469
 DB 422 TTTCTTTATCAAGTATCATTCAGCAATTAGTTGTGATTTGTTGTTCAATCATCTATC 481
 QY 470 ATACGTATTTATGATGTTGCTTACCCTGCTCCTATTTTCTGATATACACACCCGAG 529
 DB 482 ATATATTTTACTTGTGTTAGCGGAATGTTTCAATCATATTAATAATTCGACACAAT 541
 QY 530 CGAACATTAACGAATTTATCAATAAAGCAACCTAAAGTAAATGGTTATTA 582
 DB 542 CTAAATATAGTAGAATTTTAAAGGAGAGAACCTAAATTAATTAATGGATGTA 594

RESULT 12
 AA555594
 ID AA555594 standard; DNA; 642 BP.
 XX
 AC AA555594;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Streptococcus pneumoniae DNA for cellular proliferation protein #165.
 XX
 KW Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO2001/0955-A2.
 XX
 PD 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 XX Yamamoto RT, Xu HH;
 PI WPI: 2001-611495/70.
 DR P-PSDB: AAU37735.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PS Claim 27; Seq ID No 9231; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;

Query Match 17.5%; Score 101.6; DB 23; Length 642;
 Best Local Similarity 53.1%; Pred. No. 2.2e-19;
 Matches 240; Conservative 0; Mismatches 209; Indels 3; Gaps 1;
 QY 1 ATGTTAATTTGCTTTTATGATTATTTTGGCTACTTCTGATAGGACGATTCCTATCTGGCTTA 60
 DB 1 ATGATTACAATAGTATTTTATTAATCTAGCTATCTGCTGGTTCGATTCCTATCTGCTCTC 60
 QY 61 ATTTGGGCAAGCTTGGCAAGGAATTCATATTCGGGAGCAGCGGCAACTTAGGC 120
 DB 61 TGGATTGGCAAGTATTTCTTCAATCAATCTACCGAGCATGTTCTCTTAACACATGGA 120
 QY 121 GCTACCAATGCAATTCGCTACATGTTGGTGTAAAGCTGCTGCTGCTCATAGCCGAGAT 180
 DB 121 AGCACCACACCTTCCGCAATTTTAGTAAAGAGCTGTTAGGCAACCTTTGTGATTGAC 180
 QY 181 ATTTTGAAGGACACTGGCAACTGCATTTGCTTTTCTCATGCAATGTT---TGATATTCAC 237
 DB 181 TTTTCAAAAGGAACCTTAGCAACGCTGCTTCGATTTATTTTTCATACAGGCGCTTCT 240
 QY 238 CGCTTCTTCAGGAGTCTTTGCGGTTTATAGCCAGCTGTTTCCCATCTTCGCCAAATTT 297
 DB 241 CCTTTCATCTTTGGACTTTTGGCTTTATCGGCCATACCTTCCTCTATCTTCAGGATTT 300
 QY 298 AAAGCGGTAAAGCGGTGGCGACATCAGAGGCGTCTTGTCTATTTTACGACCCCTGTGA 357
 DB 301 AAAGGTGGTAGGCTGTCCCAACCAAGTCTGTGGAGTGAATTTTCGGATTTGCGCCTATCTTC 360

Streptococcus pneumoniae yneS gene encodes a polypeptide (S-yneS) essential for survival for a wide range of bacteria. Identifying an antibacterial agent comprises contacting a yneS polypeptide (S-yneS) with a test compound and detecting an interaction of the test compound with the S-yneS polypeptide which indicates that the compound is an antibacterial agent. Alternatively, detecting a decrease in function of the polypeptide contacted with the test compound and determining whether the compound inhibits growth of bacteria, relative to the growth of bacteria cultured in the absence of a test compound where inhibition of growth indicates the compound is an antibacterial agent. Inhibitors of S-yneS function are useful for treating a Streptococcus pneumoniae infection in mammals.

Sequence 642 BP; 151 A; 143 C; 122 G; 226 T; 0 other;

Query Match 17.2%; Score 100; DB 21; Length 642;

Best Local Similarity 52.9%; Pred. No. 6.5e-19;

Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 ATGTTAATGCTTTATTCATTATTTGGCTTACTTGTATAGCAGCATTCCTCCATCTGGCTTA 60
 Db 1 ATGATTACAATAGTTTATTTATCTACCTTCTGCTGGTTGATTCATCTGGCTTC 60
 Qy 61 ATTTGGGCAAGCTTGGCAAGGAATGATATTCGGGACGACGGAAGCGCAACTTAGGC 120
 Db 61 TGGATTGGCAAGATATCTTCAATCAATCTACGCGAGCATGGTTCTGGTAACACTGGA 120
 Qy 121 GCTACCAATGATTCGTTACATTTGGGTAAAGCTGGTTGGTCTCATAGCGGAGAT 180
 Db 121 ACGACCAACACCTTCCGCAATTTAGTAAAGAAAGCTGGTATGGCAACCTTTGTGATGAC 180
 Qy 181 ATTTTGAAGGACACTGGCAACTGCATTGCTTTCTCATGCTGT---TGATATTCAC 237
 Db 181 TTTTTCGAAGGACCTGACGACGCTGCTCGGATATTTTCACTACAGGCGTTTCT 240
 Qy 238 CGCTCTTTCGAGGAGTCTTTGGGTTTAGGCAAGCTGTTTCCCATCTTCGCCAAATTT 297
 Db 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTCGAGGATTT 300
 Qy 298 AAGCGGTAAAGCGTGGGACATCAGGAGCGTTTGGCTATTTTACGCCACCTGTGA 357
 Db 301 AAAGGTGGTAAGGCTGTGCGCAACAGTGTCTGGAGTGATTTTGGGATTTGCGGCTATCTTC 360
 Qy 358 TTTATCAGCATGGTTGCGGTATTTCTTCTATCTTTTATATCTTACCTAAATTTGTTTCTCTC 417
 Db 361 TGCTCTACCTTGGCATATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
 Qy 418 TCATCATGTTTAAACAGGATCTATATCTGTTAT 449
 Db 421 TCTAGTGTACAGCATCGATTGCGGCTGTTAT 452

RESULT 15

AAV52170

ID AAV52170 standard; DNA; 6171 BP.

XX

AC AAV52170;

XX

DT 23-OCT-1998 (first entry)

XX

DE Streptococcus pneumoniae genome fragment SEQ ID NO:37.

XX

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

XX

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX

OS Streptococcus pneumoniae.

XX

PN WO9818931-A2.

XX

PD 07-MAY-1998.

XX

PF 30-OCT-1997; 97WO-US19588.

XX

PR 31-OCT-1996; 96US-0029960.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

PI Kunsch CA, Rosen CA;

XX

DR WPI; 1998-272225/24.

XX

Computer-readable medium with recorded Streptococcus pneumoniae polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae

XX

Claim 1; Page 371-374; 1409pp; English.

XX

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 6171 BP; 1645 A; 1424 C; 1188 G; 1914 T; 0 other;

Query Match 17.2%; Score 100; DB 19; Length 6171;

Best Local Similarity 52.9%; Pred. No. 1.6e-18;

Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1. ATGTTAATGCTTTATTCATTATTTGGCTTACTTGTATAGCAGCATTCCTCCATCTGGCTTA 60
 Db 5149 ATGATTACAATAGTTTATTTATCTACCTATCTGCTGGTTGATTCCTCACTGGCTC 5208
 Qy 61 ATTTGGGCAAGCTTGGCAAGGAATGATATTCGGGACGACGGAAGCGCAACTTAGGC 120
 Db 5209 TGGATTGGCAAGATATCTTCAATCAATCTACGCGAGCATGGTTCTGGTAACACTGGA 5268
 Qy 121 GCTACCAATGATTCGTTACATTTGGGTAAAGCTGGTTCCGTCGTCATAGCGGAGAT 180
 Db 5269 ACGACCAACACCTTCCGCAATTTAGTAAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 5328
 Qy 181 ATTTTGAAGGACACTGGCAACTGCATTGCTTTCTCATGCTGT---TGATATTCAC 237
 Db 5329 TTTTTCGAAGGAACCTTAGCAACGCTGCTCCGATATTTTTCATCTACAAGGCGTTTCT 5388
 Qy 238 CGCTTCTTGGAGGAGTCTTTGCGGTTTATAGGCCACGCTGTTTCCCATCTTCGCCAAATTT 297
 Db 5389 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCCATACCTTCCCTATCTTTGCAGGATTT 5448
 Qy 298 AAGCGGTAAAGCGTGGCGACATCAGGAGCGGTTTGTCTATTTTACGCAACCTCTGTTA 357
 Db 5449 AAAGGTGGTAGGCTGTCCCAACACAGTCTGGAGTGATTTTTCGGATTTTGCCTATCTTC 5508
 Qy 358 TTTATCAGATGTTGCGGATTTCTTCTATCTTTTATATCTTATGACTAAATTTGTTCTCTC 417
 Db 5509 TGCTCTACCTTGGCATTTATCTCTTTTGGAGCTCTCTATCTTGGCAGATGATGATTTCACTG 5568

Search completed: December 24, 2002, 22:48:36
Job time : 459.703 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002., 22:16:24 ; Search time 3199.81 seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-4
Perfect score: 582
Sequence: 1 atgttaatgctttattgat.....ctaaagtaaaatgggtataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	49.2	8.5	330	9	AL513817
C 2	49.2	8.5	389	9	AL513693
C 3	48.4	8.3	508	14	BQ391243
4	47.8	8.2	1101	17	AL108271
5	47.4	8.1	740	17	AQ331048
C 6	47.2	8.1	329	9	AL513719

7	46.4	8.0	458	9	AL514085
C 8	46.2	7.9	477	14	BQ390789
9	45.6	7.8	379	9	AL514359
C 10	45.2	7.8	514	14	BQ521704
11	44.6	7.7	413	17	B20574
C 12	44.4	7.6	329	9	AL514627
C 13	44.4	7.6	335	9	AL513597
14	44.4	7.6	413	14	A2524410
15	44.2	7.6	413	14	BQ076552
16	44.2	7.6	414	14	BQ077621
C 17	44.2	7.6	415	10	AV657079
18	44.2	7.6	474	14	BQ076863
19	44	7.6	365	9	AL515373
20	44	7.6	900	17	A2682023
21	44	7.6	1270	13	BM456675
C 22	43.8	7.5	630	17	AG154621
23	43.8	7.5	743	13	BJ446581
24	43.6	7.5	360	14	BQ400218
C 25	43.6	7.5	533	14	BQ396711
C 26	43.6	7.5	761	17	AG152484
C 27	43.4	7.5	638	9	AL513901
C 28	43.2	7.4	454	9	AL513951
C 29	43.2	7.4	651	17	CNS03PNQ
C 30	43.2	7.4	867	17	CNS00CX5
C 31	43	7.4	313	14	BQ391420
C 32	43	7.4	338	14	C90212
33	43	7.4	341	17	CNS00606
34	42.8	7.4	300	17	CNS0076L
C 35	42.3	7.4	392	9	AL514511
C 36	42.3	7.4	423	9	AL513863
C 37	42.3	7.4	889	17	BH134799
C 38	42.3	7.4	899	17	BH154279
39	42.3	7.4	957	17	BH130614
C 40	42.8	7.4	1045	17	CNS03YE4
C 41	42.6	7.3	367	14	BQ525045
C 42	42.6	7.3	870	17	CNS06022
C 43	42.6	7.3	924	17	CNS01GHN
C 44	42.4	7.3	1074	17	CNS01082
C 45	42.2	7.3	637	17	CNS0245B

ALIGNMENTS

RESULT 1
AL513817/c
LOCUS AL513817 LTI_NFL006_PL2 Homo sapiens CDNA clone CLOBA0092E06 3
DEFINITION prime, mRNA sequence.
ACCESSION AL513817
VERSION AL513817.1 GI:12777311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 330)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..330
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA0092E06"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com

BASE COUNT 183 a 28 c 17 g 88 t 14 others
ORIGIN

Query Match 8.5% Score 49.2; DB 9; Length 330;

Best Local Similarity 47.1%; Pred.No. 0.054;
Matches 114; Conservative 10; Mismatches 118; Indels 0; Gaps 0;

OY 332 TTTTGGCTATTATACGACCCCTGTTATATATACGATGGTGGGTAATCTCATCTTTT 391
||||| ||||| : ||||| : ||| : ||| : ||||| ||||| |||||

Db 277 TTTATTTTTCYCCYTT 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 392 TATATTCGACTAAATTTGTTCTCTCATCGATGTTAAACAGGATCTATACGTTATAT 451
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

Db 217 TTTTCYTTT 158
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 452 ATAGTTTCTTGTCCATGATACGTATTATGATGTCGTACCCCTGCTCACTATTTTG 511
||| ||||| : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 157 TTTTTCYTTTCYCCYTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAGGGTTTTTTT 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 512 TGATATACAGACCGGACGAACTAAACGAATTAATCAATAAACAGACCTAAAGTAA 571
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

Db 97 TTAATAAA 38
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 572 AA 573

Db 37 AA 36

RESULT 2

AL513693/c

LOCUS AL513693 389 bp mRNA linear EST 13-FEB-2001
DEFINITION AL513693 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBA007D12 3
prime, mRNA sequence.

ACCESSION AL513693

VERSION AL513693.1 GI:12777187

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 389)

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 389

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CLOBA007D12"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"

/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com

BASE COUNT 201 a 13 c 5 g 141 t 29 others
ORIGIN

Query Match 8.5% Score 49.2; DB 9; Length 389;

Best Local Similarity 45.9%; Pred.No. 0.053;

Matches 111; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

OY 332 TTTTGGCTATTATACGACCCCTGTTATATATACGATGGTGGGTAATCTTCATCTTTT 391
||||| ||||| : ||||| : ||| : ||| : ||||| ||||| |||||

Db 327 TTTTTCYTTTTCYCCYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 268
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 392 TATATTCGACTAAATTTGTTCTCTCATCGATGTTAAACAGGATCTATACGTTATAT 451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 267 TTTTTCYTTTTCYCCYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 208
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 452 ATAGTTTCTTGTCCATGATACGTATTATGATGTCGTACCCCTGCTCACTATTTTG 511
||||| ||||| : ||| : ||| ||| ||| ||| ||| ||| ||| |||

Db 207 TTTTTCYTTTTCYCCYTTTTTTTTTTTTTTTTTTTTTTTTTTTNNANGTTTNNATTTT 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 512 TGATATACAGACCGGACGAACTAAACGAATTAATCAATAAACAGACCTAAAGTAA 571
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 147 TTAATAACAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 572 AA 573

Db 87 AA 86

RESULT 3

BO391243/c

LOCUS BO391243

DEFINITION BO391243

ACCESSION BO391243

VERSION BO391243.1

KEYWORDS EST.

SOURCE western clawed frog.

ORGANISM Silurana tropicalis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Silurana.

1 (bases 1 to 508)

NIH-XCG http://image.llnl.gov/image/html/xenopuslib_info.shtml.

National Institute of Child Health and Human Development, National

Cancer Institute, Xenopus Gene Collection

Unpublished (2002)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

CDNA Library Preparation:

CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

info@image.llnl.gov

Plate: LLAM11782 row: K column: 5

Seq primer: -21M13 forward primer (ABI).

Location/Qualifiers

1. 508

/organism="Silurana tropicalis"

/db_xref="taxon:8364"

/clone="IMAGE:5308924"

/clone_lib="NICHD XGC Emb5"

/tissue_type="gastrula"

/dev_stage="embryo, stages 10-13"

/lab_host="DH10B (phage-resistant)"

/note="vector: pCMV-SPORT6.ccd; Site_1: NotI; Site_2:

EcoRV; Cloned unidirectionally. Primer: Oligo dr. Average

insert size 2.0 kb. Constructed by Invitrogen. Note: This

is a Xenopus Gene Collection (XGC) library."

BASE COUNT 316 a 42 c 56 g 94 t

ORIGIN

[illegible]

ORIGIN

```

Query Match      8.1%; Score 47.4; DB 17; Length 740;
Best Local Similarity 41.3%; Pred. No. 0.14;
Matches 144; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 162-GGTCGTCATAGCGGAGATATTTGAAGGGGACACTGGCACTGCATTCCTCTCAT 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 280 GATTATATTCCTGAAATTTTACAAGTACCGTAGCACCAGAAGTCTCTTATGAAT 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 222 GCATGTGTATATCCCGCTCTTTCGAGGAGCTCTTTCGGGTTTTCAGCCAGCTGTTC 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 340 GCATGGAGATGTTATTTGGTTATCGAAGATTTCGCTGGGATATTTATATCTAT 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 282 CATCTTCGCCCAATTTAAAGCGGTAAGCCGTCGCGACATCAGGAGCGTTTCCTATT 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 400 TTTATTTGGGGCTGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 342 TTACGACCCCTGTTATTTATACCATGGTTCGGGTATCTTCATCTTTTATACCTGAC 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 460 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 402 TAAATTTGTTCTCTCTCATCGATGTTAACAGGGATCTATACCTGTATATATAGTTCT 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 520 TTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 462 TCTCCATGATACGATATTTATGATGTCGTTACCTGCTCCTACTATTTTT 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 580 TTTNTTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AL513719/c
LOCUS      AL513719 LTI_NFL006_PL2 329 bp mRNA linear EST 13-FEB-2001
DEFINITION prime, mRNA sequence.
ACCESSION AL513719
VERSION   AL513719.1 GI:12777213
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 329)
AUTHORS   Li,W.B., Gruber,C., Jessee,J., and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             source
            1..329
                Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CLOBA0072H01"
                /clone_lib="LTI_NFL006_PL2"
                /tissue_type="placenta"
                /note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT      186 a 13 c 3 g 60 t 67 others
ORIGIN

Query Match      8.1%; Score 47.2; DB 9; Length 329;
Best Local Similarity 37.6%; Pred. No. 0.17;
Matches 144; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 332 TTTTCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATCTTCTATCTTT 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 30 TTTTCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATCTTCTATCTTT 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 392 TATCTTGCATTAATTTGTTCTCTCTCATCGATGTTAACAGGGATCTATCTATATAT 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Matches 97; Conservative 39; Mismatches 122; Indels 0; Gaps 0;

QY 325 GGAGCGGTTTGGCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATCTTC 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 GGGKTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 395 ATCTTTTATATCTGACTAAATTTGTTCTCTCTCATCGATGTTAACAGGGATCTATATCT 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 TTTTCTATTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 445 GTTATATATAGTTTCTTGTCCATGATACGATTTTATTTGATTTGCTTACCTGCTCACT 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 TTTTCTATTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 505 ATTTTGTGATATACAGACACCGACGACATTAACGAATATATCAATAAACAGAACCT 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 WTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 58
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 565 AAAGTAAATGTTATATA 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 57 AAATAAAAAAATAAAAAA 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
AL514085
LOCUS      AL514085 LTI_NFL006_PL2 458 bp mRNA linear EST 13-FEB-2001
DEFINITION prime, mRNA sequence.
ACCESSION AL514085
VERSION   AL514085.1 GI:12777579
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 458)
AUTHORS   Li,W.B., Gruber,C., Jessee,J., and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             source
            1..458
                Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CLOBA004ZE10"
                /clone_lib="LTI_NFL006_PL2"
                /tissue_type="placenta"
                /note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT      186 a 13 c 3 g 207 t 49 others
ORIGIN

Query Match      8.0%; Score 46.4; DB 9; Length 458;
Best Local Similarity 42.6%; Pred. No. 0.26;
Matches 107; Conservative 15; Mismatches 129; Indels 0; Gaps 0;

QY 332 TTTTCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATCTTCTATCTTT 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 30 TTTTCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATCTTCTATCTTT 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 392 TATCTTGCATTAATTTGTTCTCTCTCATCGATGTTAACAGGGATCTATCTATATAT 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 90 TTTTTCCTTCCTCCATGATCGTATTATTCATTGCTGTACCTCGTCACTATTTCCTT 149
Qy 452 ATAGTTTCCTTCCTCCATGATCGTATTATTCATTGCTGTACCTCGTCACTATTTCCTT 511
Db 150 WTTWTATWTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 209
Qy 512 TGATATACAGACCGAGCGAACATTAACGAATATCAATAAAAAACAGAACCTAAAGTAA 571
Db 210 NTTTTTTCCTCCACCTGCTTTAAARAAAAAARAAAAAARAAAAAARAAAAAAR 269
Qy 572 AATGGTTATAA 582
Db 270 AAARAAAAA 280

RESULT 8
BQ390789/c
LOCUS BQ390789 477 bp mRNA linear EST 22-MAY-2002
DEFINITION NISC_mgl5003.x1 NICHD XGC Emb5 Silurana tropicalis cDNA clone
IMAGE:5308372 3', mRNA sequence.
ACCESSION BQ390789
VERSION BQ390789
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 477)
AUTHORS NIH-XGC http://image.llnl.gov/image/html/xenopuslib\_info.shtml.
TITLE National Institute of Child Health and Human Development, National
JOURNAL Cancer Institute, Xenopus Gene Collection
COMMENT Unpublished (2002)
Contact: Robert Strausberg, Ph. D.
Email: cgapbs-remail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM11781 row: D column: 5
Seq primer: -21M13 forward primer (ABI).
FEATURES
source
1..477
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="IMAGE:5308372"
/tissue_type="gastrula"
/dev_stage="embryo, stages 10-13"
/lab_host="DH10B (phage-resistant)"
/notes="Vector: pCMV-SPORT6.ccd; Site_1: NotI; Site_2:
EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average
insert size 2.0 kb. Constructed by Invitrogen. Note: This
is a Xenopus Gene Collection (XGC) library."
BASE COUNT 245 a 18 c 15 g 199 t
ORIGIN
Query Match 7.9%; Score 46.2; DB 14; Length 477;
Best Local Similarity 49.4%; Pred. No. 0.29;
Matches 120; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 331 GTTTGCTATTTACACACCCCTGTTATTTATTCACGATGGTTGCGGTATTCCTCACTTT 390
Db 346 GTTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 287
Qy 391 TTATCTTCCTCAATTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 450
Db 286 TTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 227
Qy 451 TATAGTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 510

```

```

Db 226 TTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 167
Qy 511 GTGATATACACACCGAGCGAACATTAACGAATATCAATAAAAAACAGAACCTAAAGTAA 570
Db 166 TAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 107
Qy 571 AAA 573
Db 106 AAA 104

RESULT 9
AL514359
LOCUS AL514359 379 bp mRNA linear EST 13-FEB-2001
DEFINITION AL514359 LFI_NFL006_PL2 Homo sapiens cDNA clone CL0BB0112B11 3
prime, mRNA sequence.
ACCESSION AL514359
VERSION AL514359.1 GI:12777853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 379)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BB0112B11"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; Site_2: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 134 a 5 c 20 g 186 t 34 others
ORIGIN
Query Match 7.8%; Score 45.6; DB 9; Length 379;
Best Local Similarity 46.9%; Pred. No. 0.42;
Matches 105; Conservative 10; Mismatches 109; Indels 0; Gaps 0;

Qy 353 TGTATTTATACAGATGTTGCGGTATTCCTCACTTCCTTATACCTGACTAAATTCGTTT 412
Db 15 TTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 74
Qy 413 CTCCTCTCGATGTTAAACAGGATCTATACCTGTTATATATATAGTTTCCTTGCCATGATA 472
Db 75 TTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 134
Qy 473 CGTATTTATTCGTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 532
Db 135 TTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 194
Qy 533 ACAATTAACGAATTAACAATAAACAACCTAAAGTAAAGG 576
Db 195 WAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 238

```

RESULT 10

B0521704/c
 LOCUS B0521704 514 bp mRNA linear EST 10-JUN-2002
 DEFINITION NISC_nll2ell.x1 NICHDXGC_Emb7 Silurana tropicalis cDNA clone
 IMAGE:5336181.3, mRNA sequence.

ACCESSION

B0521704
 B0521704.1 GI:21380583

KEYWORDS

EST.

SOURCE

western clawed frog.

ORGANISM

Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 xenopodinae; Silurana.
 1 (bases 1 to 514)
 NIH-XGC http://image.llnl.gov/image/html/xenopuslib_info.shtml.
 National Institute of Child Health and Human Development, National
 Cancer Institute, Xenopus Gene Collection
 Unpublished (2002)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

JOURNAL

COMMENT
 cDNA Library Preparation:
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC)
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 info@image.llnl.gov
 Plate: LAM11853 row: J column: 22
 Seq primer: -21M13 forward primer (ABI).

FEATURES

source
 1..514
 /organism="Silurana tropicalis"
 /db_xref="taxon:8364"
 /clone="IMAGE:5336181"
 /clone_lib="NICHDXGC_Emb7"
 /tissue_type="tailbud"
 /dev_stage="embryo, stages 20-27"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pCMV-SPORT6.cdb; Site_1: NotI; Site_2:
 EcoRV; Cloned unidirectionally. Primer: Oligo dt. Average
 insert size 2.1 kb. Constructed by Invitrogen. Note: This
 is a Xenopus Gene Collection (XGC) library."

BASE COUNT

341 a 9 c 5 g 159 t

ORIGIN

Query Match 7.8%; Score 45.2; DB 14; Length 514;
 Best Local Similarity 49.2%; Pred. No. 0.51;
 Matches 119; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 332 TTTTCGCTATTTACGCACCCCGTATTATATACAGATGGTTCGGTATTCATCTTTT 391

Db 335 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 276

Qy 392 TATACCTGACTAAATTTGTTCTCTCATCGATGTTAACAGGGATCTATCTGTATAT 451

Db 275 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 216

Qy 452 ATAGTTTCTTGTCCATGATAGTATTATGATTGTCGTACCGTCTACTATTTTG 511

Db 215 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 156

Qy 512 TCATATACAGACACCGAGCAATTAACGAATTTATCAATAAACAACCACTTAAGTAA 571

Db 155 TTTTATATAA 96

Qy 572 AA 573

Db 95 AA 94

RESULT 11

B05074
 LOCUS B05074 413 bp DNA linear GSS 16-SEP-1997

DEFINITION

T20L8-T7 TAMU Arabidopsis thaliana genomic clone T20L8, DNA
 sequence.

ACCESSION

B20574
 B20574.1 GI:2395628

KEYWORDS

GSS.

SOURCE

thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 413)
 Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
 Ecker, J.

TITLE

BAC End Sequences at ATGC

JOURNAL

Unpublished (1997)

COMMENT

Other_GSSs: T20L8-Sp6
 Contact: Ecker, J.
 Arabidopsis Thaliana Genome Center
 University of Pennsylvania
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA
 19104

Tei: 215-898-9384

Fax: 215-898-8780

Email: jecker@atgenome.bio.upenn.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 95

High quality sequence stop: 103.

FEATURES

source
 1..413
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T20L8"
 /clone_lib="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelOBAII; Site_1: HindIII; Site_2: HindIII
 ; Produced by Rod Wing"

BASE COUNT

20 a 68 c 24 g 293 t 8 others

ORIGIN

Query Match 7.7%; Score 44.6; DB 17; Length 413;
 Best Local Similarity 53.1%; Pred. No. 0.75;
 Matches 93; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 332 TTTTGCATTTTACGCACCCCGTATTATATATACAGATGGTTCGGTATTCATCTTTT 391

Db 234 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 293

Qy 392 TATACCTGACTAAATTTGTTCTCTCATCGATGTTAACAGGGATCTATCTTATAT 451

Db 294 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 353

Qy 452 ATAGTTTCTTGTCCATGATAGTATTATATGATTGTCGTACCGTCTACTATTTT 510

Db 354 CTCTCTCTTTTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 412

RESULT 12

AL514627/c
 LOCUS AL514627 329 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL514627 LTL_NFL006_PL2 Homo sapiens cDNA clone CL08B002E11 3
 prime, mRNA sequence.

ACCESSION

AL514627

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 329)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:25:54 ; Search time 84.875 Seconds
(without alignments)
2102.926 Million cell updates/sec

Title: US-10-068-080-4

Perfect score: 582
Sequence: 1 atgttaattgctttattgat.....ctaaagtaaaatggtttataa 582

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	582	100.0	582	4	US-09-222-938A-71
2	108.2	18.6	615	4	US-09-134-001C-2402
3	100	17.2	6171	4	US-08-961-527-37
4	76.6	13.2	344	4	US-09-222-938A-14
5	48.8	8.4	7218	1	US-08-232-463-14
6	37	6.4	19124	2	US-08-487-826B-13
7	35.8	6.2	1723	4	US-09-181-336-14
8	34.2	5.9	1320	1	US-08-257-073-15
9	34.2	5.9	1482	4	US-08-098-327E-41
10	34.2	5.9	1482	4	US-08-462-625-41
11	34.2	5.9	1493	4	US-08-098-327E-38
12	34.2	5.9	1493	4	US-08-462-625-38
13	34	5.8	950	4	US-08-098-327E-32
14	34	5.8	950	4	US-08-462-625-32
15	33.8	5.8	152331	3	US-09-128-155-16
16	33.8	5.8	176373	3	US-09-128-155-17
17	33.4	5.7	375	4	US-09-134-001C-2507
18	33	5.7	3275	4	US-09-370-838-151
19	32.8	5.6	473	1	US-08-764-100-16
20	32.8	5.6	4970	1	US-08-764-100-14
21	32.8	5.6	4970	1	US-08-764-100-20
22	32.6	5.6	1482	4	US-08-098-327E-45
23	32.6	5.6	1482	4	US-08-462-625-45
24	32.4	5.6	15611	4	US-09-441-340-1
25	32	5.5	994	4	US-09-018-584A-42
26	32	5.5	2178	4	US-09-334-601-6
27	32	5.5	2288	4	US-09-334-601-1

28	32	5.5	2448	1	US-08-526-964-2	Sequence 2, Appli
29	32	5.5	2448	2	US-08-946-617-2	Sequence 2, Appli
30	32	5.5	2448	3	US-09-031-897-2	Sequence 2, Appli
31	32	5.5	5203	4	US-09-257-770-1	Sequence 1, Appli
C 32	31.8	5.5	262	2	US-08-867-030B-3	Sequence 3, Appli
C 33	31.8	5.5	262	5	PCT-US95-06119-3	Sequence 3, Appli
34	31.4	5.4	318	4	US-09-134-001C-2112	Sequence 3, Appli
C 35	31.4	5.4	1956	4	US-08-559-896B-1	Sequence 2112, Ap
C 36	31.4	5.4	2187	4	US-09-134-001C-2131	Sequence 1, Appli
37	31.4	5.4	2293	4	US-09-336-643A-5	Sequence 2131, Ap
38	31.2	5.4	1338	4	US-08-887-534A-60	Sequence 5, Appli
C 39	31.2	5.4	3494	4	US-09-334-601-5	Sequence 60, Appli
40	31.2	5.4	6124	4	US-08-213-419B-3	Sequence 5, Appli
41	31	5.3	427	4	US-09-397-787-209	Sequence 3, Appli
C 42	31	5.3	2013	4	US-09-134-001C-103	Sequence 103, App
C 43	31	5.3	2447	2	US-09-014-969-14	Sequence 103, App
44	31	5.3	2984	4	US-09-605-785-335	Sequence 14, Appli
45	31	5.3	2984	4	US-09-439-313-335	Sequence 335, App

ALIGNMENTS

RESULT 1
US-09-222-938A-71
; Sequence 71, Application US/09222938A
; Patent No. 5437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Chrisian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 71:
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(579)
US-09-222-938A-71

Query Match	100.0%	Score 582;	DB 4;	Length 582;
Best Local Similarity	100.0%;	Pred. No. 1e-169;		
Matches 582;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTTAATGCTTTATTGATTATTTGGCCTACTTGTATAGGACGATTCATCTGGCTTA	60	
DB	1	ATGTTAATGCTTTATTGATTATTTGGCCTACTTGTATAGGACGATTCATCTGGCTTA	60	
QY	61	ATTGTGGGCAAGCTTCCCAAGGAATTGATTCGGGAGCACCAGCGCAACTTAGGC	120	
DB	61	ATTGTGGGCAAGCTTCCCAAGGAATTGATTCGGGAGCACCAGCGCAACTTAGGC	120	
QY	121	GCTACCAATGCAATCCGTCATACATTTGGGTGTTAAAGCTGGTTCGTCATAGCCGGAGAT	180	
DB	121	GCTACCAATGCAATCCGTCATACATTTGGGTGTTAAAGCTGGTTCGTCATAGCCGGAGAT	180	
QY	181	ATTTTGAAGGGACACTGGCACTGCCTTTTCTCATGATGTTGATATTCACCCG	240	
DB	181	ATTTTGAAGGGACACTGGCACTGCCTTTTCTCATGATGTTGATATTCACCCG	240	
QY	241	CTTCTTGCAGAGCTCTTGGGGTTTGTAGCCACGCTGTTCCCATCTTCGCCAAATTTAA	300	
DB	241	CTTCTTGCAGAGCTCTTGGGGTTTGTAGCCACGCTGTTCCCATCTTCGCCAAATTTAA	300	
QY	301	GGCGTAAAGCCCGTGGCGACATCAGGAGCGTTTTCGTATTTTACGACCCCTGTTATT	360	
DB	301	GGCGTAAAGCCCGTGGCGACATCAGGAGCGTTTTCGTATTTTACGACCCCTGTTATT	360	

Db 301 GCGGTAAGCGGTGGCGACATCAGGAGCGGTTTGTCTATTTTACGCACCCCTGTTATTT 360
Qy 361 ATACAGATGTTGGGTATTTCTTCATCTTTTATACCTTTTATCTGTTGTTCTCTCTCA 420
Db 361 ATACAGATGTTGGGTATTTCTTCATCTTTTATACCTTTTATCTGTTGTTCTCTCTCA 420
Qy 421 TCGATGTTAAACAGGATCTATACCTGTTATATATATAGTTCTTTTGTCCATGATAGTATTTA 480
Db 421 TCGATGTTAAACAGGATCTATACCTGTTATATATATAGTTCTTTTGTCCATGATAGTATTTA 480
Qy 481 TTGATTCGTTACCCCTGCTCCTACTATTTTGTGATATACAGACCGAGGAAACATTTAA 540
Db 481 TTGATTCGTTACCCCTGCTCCTACTATTTTGTGATATACAGACCGAGGAAACATTTAA 540
Qy 541 CGAATTATCAATAAACAAGAACCTAAAGTAAATGGTTATAA 582
Db 541 CGAATTATCAATAAACAAGAACCTAAAGTAAATGGTTATAA 582

RESULT 2

US-09-134-001C-2402
; Sequence 2402, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2402
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2402

Query Match 18.6%; Score 108.2; DB 4; Length 615;
Best Local Similarity 50.7%; Pred. No. 6.7e-24;
Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

Qy 1 ATGTTAATGCTTTATTCATATTTTGGCTTACTTGTATAGGACGACATTCATCTGCTTA 60
Db 7 ATGATGATCATCGTCATCTTAATCTTGATGTTATCTGATTCGTCATTCCTCAAGCGGTTA 66
Qy 61 ATTGTGGCAAGCTTGCCAAAGGAATGATATTCGGGAGCAGCAGGCAAGCACTTAGGC 120
Db 67 ATATTGGTAAATTTATTTTAAAGATATAAGACATACGTTAGTGAATACCTGGA 126
Qy 121 GCTACCAATGCATTCCTGATATGTTGGTGTAAAGCTGTTTGGTGTGTCATAGCGGAGAT 180
Db 127 GCAACTAACAGTTTCTGTTCTTGGGAAGACCACTGGATTTATAGTTACGTTTATAGAT 186
Qy 181 ATTTTGAAGGGACATGGGCACTGCATTCGCTTTTCTCATGCATGTTGATATTCACCG 240
Db 187 ATTTTGAAGGGATTTATACAGTCTTTTTCACACTGTTTCCAGATTCATCGGATGGT 246
Qy 241 CTCTCTT-----GCAGGAGTCTTTGCGGTTTTAGGCGCAC 273
Db 247 GTTATAGCACCTTCTTTTACAAATGGTTTATAGTAGATGTTTGTCAATCTCGGTCAC 306
Qy 274 GTGTTCCCATCTTCGCAAAATTTAAAGCGGTTAAAGCCGTGGCGACATCAGGAGCGGTT 333
Db 307 GTGATCAATATATCTGAAATTTAATGGCGGAAAGACAGTACGTACCACTGCGAGGAT 366
Qy 334 TTGCTATTTTACGACCCCTGTTATTTATACAGATGTTGCGGTTATCTTCATCTTTTAA 393
Db 367 GTATTAGTGTCAAATCCTATTTTACTTCTTATCTTGGCAATATCTTTTATAGTGATTA 426

Qy 394 TACTTGACTAAATTTGTTCTCTCTCATCGATGTTTAAACAGGATCTATCTGTTATATAT 453
Db 427 AATATCTTTAAATATATGTTCTTTTATCAAGTATCATTTGCAGCAATTTAGTTGTGATGGT 486
Qy 454 AGTTCTTTTGCATGATAGTATTTATTTGATTTGCTGTTACCTGCTCACTATTTTGTG 513
Db 487 TCAATCATCATTTCTGATTTATTTTACTTGTCTGTTAGCGGAATTTGTTTCAATCATATTA 546
Qy 514 ATATACAGACCGGCGGACATTAACGAATTTATCAATAAACAAGCACTTAAAGTAA 573
Db 547 ATAAATCGACACAATTAATAATAGTTAGAAATTTTAAAGGAGGAAGCACTTAAATTA 606
Qy 574 TGGTTATAA 582
Db 607 TGCATGTAA 615

RESULT 3

US-08-961-527-37
; Sequence 37, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-37

Query Match 17.2%; Score 100; DB 4; Length 6171;
Best Local Similarity 52.9%; Pred. No. 6.9e-21;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 ATGTTAATGCTTTATTTGATTTTGGCTTACTTGTATAGGACGACATTCATCTGCTTA 60
Db 5149 ATGATTACAATAGTATTTTATTAATCTAGCCTATCTGCTGGTTCGATTCATCTGCTCTC 5208
Qy 61 ATTCTGGCGAAGCTTCCCAAGGAATTTGATTTCCGGAGCAGCAGGCAAGCACTTAGGC 120
Db 5209 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCGAGCATGTTCTTGGTAACTGGA 5268
Qy 121 GCTACCAATCATTCCTGATACATTTGGGTGTAAGAGCTGGTTCGTCGTCATAGCGGAGAT 180


```

Db 1293 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1352
Qy 489 CGTACCTGCTCACTATTTT 510
Db 1353 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1374

RESULT 6
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match 6.4%; Score 37; DB 2; Length 19124;
Best Local Similarity 57.3%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;
Matches 67; Conservative 0;

Qy 377 TATCTTCATCTTTATACCTGACTAAATTTGTTCTCTCATCGATGTTAAACAGGA 436
Db 15907 TTTTAAATTTTATTTATTAATTAATTTTATTTATTTATTTTATTTATTTAT 15848

Qy 437 TCATACGTATATATAGTTCTTGTCCATGATACGATATTTATGATTCGTTA 493
Db 15847 TAAATTTTATTTATTTATTTTATTTTAAATAAATTTTATTTATTTATGTA 15791

RESULT 7
US-09-181-336-14
; Sequence 14, Application US/09181336A
; Patent No. 6242238

```

```

; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HADLETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(1647)
; US-09-181-336-14

Query Match 6.2%; Score 35.8; DB 4; Length 1723;
Best Local Similarity 51.6%; Pred. No. 0.21; Mismatches 0; Indels 0; Gaps 0;
Matches 82; Conservative 0;

Qy 306 TAAAGCGGTGGCGACATCAGAGCGTGTTCGCTATTTTACGACACCCCTGTTATTATCAC 365
Db 1490 TAAGACCTTTGGGACCTCATGGATTACTTCCAAATCTGTCCAACTCAATGGTCTAATC 1549

Qy 366 GATGTTGGCGTATTTCTTCATCTTTTATACCTTGACTTAAATTTGTTCTCTCATCGAT 425
Db 1550 TAAAGATGTGGATGATCAACCTTCCACCTTTAATGGAACACCTCTCCGCCAGGA 1609

Qy 426 GTTACAGGATCTATCTGTTATATATAGTTCTTTTGT 464
Db 1610 GTTCACTGGTTCGCCAGCTTCTCTCATATAGTTTGT 1648

RESULT 8
US-08-257-073-15/C
; Sequence 15, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305

```

FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-15

Query Match: 5.9%; Score 34.2; DB 1; Length 1320;
Best Local Similarity 48.7%; Pred. No. 0.58;
Matches 93; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
Qy 332 TTTTCTATTATTTACGACCCCTGTTATTATCATCAGATGGTTGCGGTATCTTCATCTTTT 391
Db 812 TTTTCTATGTCAGCAGATTATCTCTTTTCTGCTGTTTATTGAACCATCTTCTTT 753
Qy 392 TATATCTGACTAAATTTGTTCTCTCATCGATGTTAACAGGGATCTATCTGTTAT 451
Db 752 TCTTCAAGATGTCCTTATGTATATCTCTGCTTCAACATTTGTTGTAATAGATTCT 693
Qy 452 ATAGTTTCTTCTCCATGATAGTATTATTCATGTTCTGTTACCCCTGCTCACTATTTTG 511
Db 692 CTATTTGTTTTTCTATATAGATATTTCCTTGGATCTCTACTATTCCTCTGTTGCC 633
Qy 512 TGATATACAGA 522
Db 632 TGAGGTAAAGA 622

RESULT 9
US-08-098-327E-41/c
Sequence 41, Application US/08098327E
Patent No. 6270771
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 1482 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1482
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-41

Query Match: 5.9%; Score 34.2; DB 4; Length 1482;
Best Local Similarity 48.7%; Pred. No. 0.61;
Matches 93; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
Qy 332 TTTTGTATTATTTACGACCCCTGTTATTATCATCAGATGGTTGCGGTATCTTCATCTTTT 391
Db 974 TTTTGTATGTCAGCAGATTATCTCTTTTCTGTTGTTTATTGAACCATCTTCTTT 915
Qy 392 TATATCTGACTAAATTTGTTCTCTCATCGATGTTAACAGGGATCTATCTGTTAT 451
Db 914 TCTTCAAGATGTCCTTATGTATATCTCTGCTTCAACATTTGTTGTAATAGATTCT 855
Qy 452 ATAGTTTCTTCTCCATGATAGTATTATTCATGTTGTTACCCCTGCTCACTATTTTG 511
Db 854 CTATTTGTTTTTCTATATAGATATTTCCTTGGATCTCTACTATTCCTCTCTGTTGCC 795
Qy 512 TGATATACAGA 522
Db 794 TGAGGTAAAGA 784

RESULT 10
US-08-462-625-41/c
Sequence 41, Application US/08462625
Patent No. 6319502
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286

Query Match 5.8%; Score 34; DB 4; Length 950;
Best Local Similarity 50.0%; Pred. No. 0.57;

RESULT 14
US-08-462-625-32/c

APPLICANT: GUERIN-MARCHAND, Claudine

; COMPUTER READABLE FORM:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 91 01286
 FILING DATE: 05-FEB-1991

NAME: MCGOWAN, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 950 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-32

Query Match 5.8%; Score 34; DB 4; Length 950;
Best Local Similarity 50.0%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 353 TGTATTATACAGATGTTGGGATTTCTTCATCTTTTATACATGACTAAATTTGTTT 412
DB 943 TCTTCTTTTGTCTGTTTATTGAACCATCTTCTTCTCAAGATGCTCTTATGT 884
QY 413 CTCTCTCATCGATGTTAAACAGGATCTATACCTGTTATATATAGTTTCTTTGTCATGATA 472
DB 883 ATATCCCTTCGTCCTTCAACATTTGTTGTAATAGATTCCTATTTGTTTTCTATTATA 824
QY 473 CGTATTATTATGTTGCTTACCCTGCTCCTATTTTCTGATATACAGA 522
DB 823 GATATTTCCTGGAATCTCTACTATTTCCCTCTGTTGCTGAGGTAAGA 774

RESULT 15
US-09-128-155-16/C
Sequence 16, Application US/09128155
Patent No. 6117634
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 5.8%; Score 33.8; DB 3; Length 152331;
Best Local Similarity 49.7%; Pred. No. 7.7;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 390 TTTATACCTGACTAAATTTGTTCTCTCATCGATGTTAAACAGGATCTATACGTAT 449
DB 37151 TTCTTCTTACCTAAATTTATTTCTTTAAGCTTTCTTACCAAAAAATACCACCTTAT 37092
QY 450 ATATAGTTTCTTTGTCATGATACGATTTATTGATTCGTTACCTGCTCCTACTATTT 509
DB 37091 TCTATAATTTCTTTACATTTCCCTTATTTCTCTGTTCTTACCTTTGTTTATACATA 37032
QY 510 TGTGATATACACACCGAGCGAATTAACGAAATTAATCAATAAAACAGAAC 562
DB 37031 ACTTTTCAATAAACTTGAATTAGACAAAAATTTTACCTTTTAAAAAGGAC 36979

Search completed: December 25, 2002, 14:26:19
Job time : 135.875 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 ; Search time 171.89 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-4
perfect score: 582
Sequence: 1 atgttaattgcctttattgat.....ctaaataaaaaatcggttataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues
Total number of hits satisfying chosen parameters: 726948

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	582	100.0	582	12	US-10-068-080-4	Sequence 4, Appli
2	109.6	18.8	6691	10	US-09-070-927A-88	Sequence 88, Appl
3	101.6	17.5	642	10	US-09-815-242-9321	Sequence 9231, Ap
4	101.6	17.2	642	10	US-09-815-242-9438	Sequence 9438, Ap
5	100	17.2	642	12	US-10-068-080-2	Sequence 2, Appli
6	90.6	15.6	609	10	US-09-823-246-1	Sequence 1, Appli
7	90.6	15.6	609	10	US-09-815-242-8039	Sequence 8039, Ap
8	88.2	15.2	606	10	US-09-815-242-4250	Sequence 4250, Ap
9	48	8.2	600	10	US-09-815-242-6900	Sequence 6900, Ap
10	46.4	8.0	72	10	US-09-874-300-4100	Sequence 4100, Ap
11	43	7.4	789	10	US-09-815-242-7356	Sequence 7356, Ap
12	42.4	7.3	612	10	US-09-815-242-9659	Sequence 9659, Ap
13	40.2	6.9	449	10	US-09-360-352-2412	Sequence 2412, Ap
14	39.8	6.8	663	10	US-09-815-242-7517	Sequence 7517, Ap
15	38	6.5	397658	10	US-09-813-320-3	Sequence 3, Appli
16	37.8	6.5	618	10	US-09-815-242-6206	Sequence 6206, Ap
17	37.2	6.4	6786	9	US-10-010-066-72	Sequence 72, Appl
18	36.6	6.3	300	10	US-09-360-352-6251	Sequence 6251, Ap
19	36.6	6.3	312	10	US-09-360-352-8414	Sequence 8414, Ap

ALIGNMENTS

```

RESULT 1
US-10-068-080-4
; Sequence 4, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, E
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/07
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Vers
; SEQ ID NO 4
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Streptococcus pneumonia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(579)

```

20	36.6	6.3	70768	9	US-10-135-322-13	Sequence 13, Appl
C 21	35.8	6.2	425	10	US-09-834-975-451	Sequence 451, Appl
C 22	35.6	6.1	424	10	US-09-960-352-11218	Sequence 11218, A
C 23	35.4	6.1	375	10	US-09-960-352-15014	Sequence 15014, A
C 24	35.2	6.0	453	10	US-09-960-352-7418	Sequence 7418, Ap
25	35.2	6.0	774	10	US-09-910-943-506	Sequence 506, App
26	35.2	6.0	99916	10	US-09-816-095-3	Sequence 3, Appl
27	34.8	6.0	397	10	US-09-960-352-13784	Sequence 13784, A
28	34.8	6.0	523	10	US-09-864-761-30606	Sequence 30606, A
29	34.6	5.9	414	10	US-09-960-352-6528	Sequence 6528, Ap
30	34.6	5.9	1503841	9	US-09-946-807-1	Sequence 1, Appl
31	34.6	5.9	1503841	10	US-09-795-668-1	Sequence 1, Appl
32	34.6	5.9	1503841	10	US-09-795-668-1	Sequence 1, Appl
33	34.4	5.9	415	10	US-09-960-352-2223	Sequence 2223, Ap
C 34	34.4	5.9	1198	10	US-09-789-561-55	Sequence 55, Appl
35	34.4	5.9	66686	10	US-09-736-960-86	Sequence 86, Appl
36	34.2	5.9	598	10	US-09-864-761-14959	Sequence 14959, A
37	34.2	5.9	598	10	US-09-864-761-31457	Sequence 31457, A
38	34.2	5.9	995	10	US-09-070-927A-833	Sequence 833, App
C 39	34.2	5.9	1482	10	US-09-837-344-41	Sequence 41, Appl
C 40	34.2	5.9	1493	10	US-09-837-344-38	Sequence 38, Appl
41	34.2	5.9	1580	10	US-09-070-927A-347	Sequence 347, App
42	34.2	5.9	1920	10	US-09-800-065-1	Sequence 1, Appl
43	34.2	5.9	5513	10	US-09-800-065-3	Sequence 3, Appl
C 44	34.2	5.9	21724	10	US-09-764-864-1603	Sequence 1603, Ap
45	34.2	5.9	21727	10	US-09-764-864-1604	Sequence 1604, Ap

Db 181 ATTTGAAAGGACACTGGCAACTGCATTGCTTTCTCATGCATGTTGATATTCACCCG 240
Qy 241 CTCTTCGACAGCTCTTTGCGGTTTAGGCCACCTGTTCCCATCTTCGCCAAATTTAAA 300
Db 241 CTCTTCGACAGCTCTTTGCGGTTTAGGCCACCTGTTCCCATCTTCGCCAAATTTAAA 300
Qy 301 GCGGTAAAGCGTGGCGACATCAGGAGCGCTTTTGTATTTTACGACACCCCTCTTATT 360
Db 301 GCGGTAAAGCGTGGCGACATCAGGAGCGCTTTTGTATTTTACGACACCCCTCTTATT 360
Qy 361 ATCAGATGTTGGGGTATCTTCATCTTTTATATATAGTTTCTTTGTCCTATTTA 420
Db 361 ATCAGATGTTGGGGTATCTTCATCTTTTATATATAGTTTCTTTGTCCTATTTA 420
Qy 421 TCGATGTTAAAGGATCTATCTGTTATATATAGTTTCTTTGTCCTATTTA 480
Db 421 TCGATGTTAAAGGATCTATCTGTTATATATAGTTTCTTTGTCCTATTTA 480
Qy 481 TTGATGTCGTTACCTCTCCTCCTATTTTGTGATATACACACCGAGCGCAATTTAAA 540
Db 481 TTGATGTCGTTACCTCTCCTCCTATTTTGTGATATACACACCGAGCGCAATTTAAA 540
Qy 541 CGAATTATCAATAAAGACACCTAAAGTAAATGGTTTATAA 582
Db 541 CGAATTATCAATAAAGACACCTAAAGTAAATGGTTTATAA 582

RESULT 2
US-09-070-927A-88/c
; Sequence 88, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Steven Barash
; Patrick J. Dillon
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6691 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-070-927A-88
Query Match 18.8%; Score 109.6; DB 10; Length 6691;
Best Local Similarity 53.4%; Pred. No. 8.9e-21;
Matches 254; Conservative 0; Mismatches 219; Indels 3; Gaps 1;
Qy 1 ATGTTAATGCTTATTGATTATTTTGGCTTACTTGTATGATGACGACATCTCCATCTGGCTTA 60
Db 5424 ATGAAATCGTCATTTTGTACTTGTTCCTTATTTATAGTTCGATTCCTCAGGTGTT 5365
Qy 61 ATTGTGGCAAGCTTCCAAAGGAATTTGATATTCGGGACGACGGAAGCGCAACTTAGGC 120
Db 5364 TGGATTGGTAAACCTTTCTTTTAAAGAGATATACGCCAATTTGGGAGTGGGAATACAGGA 5305
Qy 121 GCTACCAATGATCTCCATACATTTGGGTGTTAAAGCTGGTTCGGTCGTCATAGCCGGAGAT 180
Db 5304 ACACCAATGATCTCCATACATTTGGGTGTTAAAGCTGGTTCGGTCGTCATAGCCGGAGAT 5245
Qy 181 ATTTTGAAGGACACTGGCAACTGCATTGCTTTTCTCATGCATGT--TGATATTTCAC 237
Db 5244 ATCTGAAGAGAACTAGGCACCTTCAATACCTTATTTGTTGTTTACAAAGCGTGAAT 5185
Qy 238 CCGCTTCTTCAGGAGTCTTTGCGGTTTATAGGCCACGTTTCCCATCTTCGCCAAATTT 297
Db 5184 CCGCTTCTTCAGGAGTCTTTGCGGTTTATAGGCCACGTTTCCCATCTTCGCCAAATTT 5125
Qy 298 AAGGCGGTAAAGCGTGGCGACATCAGGAGCGCTTTGCTATTTTATAGCACCCCTGTTA 357
Db 5124 AAGGCGGTAAAGCGTGGCGACATCAGGAGCGCTTTGCTATTTTATAGCACCCCTGTTA 5065
Qy 358 TTTATCAGATGTTGCGGTATTTCTTCATCTTTTATATGTTTACATAAATTTGTTCTCTC 417
Db 5064 TTTATTTATCTGCTCTTTATTTTGTGATTTGCTGATCTATCTAACCAATGATGGTGA 5005
Qy 418 TCATCATGTTAAAGGAGTCTATATCTGTTATATATATAGTTTCTTGTCCCATGATAC 473
Db 5004 ACAAGTATGATTAGCGCTGTACTAATTAACACTTCTTACTATATTTTGGCTTTTCAC 4949

RESULT 3
US-09-815-242-9231
; Sequence 9231, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9231
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(642)
US-09-815-242-9231

Query Match 17.5%; Score 101.6; DB 10; Length 642;
Best Local Similarity 53.1%; Pred. No. 5.6e-19;
Matches 240; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

QY 1 ATGTTAAATGCTTTATTGATTTATTTGGCTACTTGCATAGGAGCAGCATTCATCTGGCTTA 60
DB 1 ATGTTAAATGCTTTATTGATTTATTTGGCTACTTGCATAGGAGCAGCATTCATCTGGCTTC 60
QY 61 ATTGTGGGCAAGCTTCCCAAGGAATTTGATTTCCGGGAGCAGGAAAGCGCAACTTTAGGC 120
DB 61 TGGATTGGACAGTATTTCTTCAATCAATCTACGCGAGCATGGTTCTGGTAACACTGGA 120
QY 121 GTTACCAATGCAATTCGATATTTGGGTGTAAGAGCTGCTGCTGCTATAGCCGGAGAT 180
DB 121 ACACCAACACCTTCGCGATTTTAGGTAAGAAAGCTGGTATGCAACCTTTTGTGATTGAC 180
QY 181 ATTTTGAAGGACACTGGCAACTGCATTCCTTTTCTCATGCATCT---TGATATTAC 237
DB 181 TTTTTCAAAGGAACCTTAGCAAGCTGCTTCGGATTTATTTTTCATCAAGCGGTTTCT 240
QY 238 CCCTTCTTTCGAGAGCTTTTGGGTTTATAGCCAGCTGTTTCCCATCTTCGCGCAAAATTT 297
DB 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCGCATACCTTCCCTATCTTTCGAGATT 300
QY 298 AAAGCGGTAAAGCGTGGGACATCAGGAGCGGTTTTCGATATTTTACGACCCCTGTTA 357
DB 301 AAAGGTGTAAGGCTGTCGCAACAGTGTGAGAGTATTTTCGGATTTTCGCGCTATCTTC 360
QY 358 TTATACAGATGTTGGGTATTTCTCATCTTTTATATCTGACTAAATTTGTTCTCTC 417
DB 361 TGTCTTACCTTCGGATTAFTCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
QY 418 TCATCATGTTAAAGGATCTATACCTGTTAT 449
DB 421 TCTAGTGTACACCATCGATCGCGGCTGTTAT 452

RESULT 4
US-09-815-242-9438
; Sequence 9438, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9438
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(642)
US-09-815-242-9438

Query Match 17.5%; Score 101.6; DB 10; Length 642;
Best Local Similarity 53.1%; Pred. No. 5.6e-19;
Matches 240; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

QY 1 ATGTTAAATGCTTTATTGATTTATTTGGCTACTTGCATAGGAGCAGCATTCATCTGGCTTA 60
DB 1 ATGTTAAATGCTTTATTGATTTATTTGGCTACTTGCATAGGAGCAGCATTCATCTGGCTTC 60
QY 61 ATTGTGGGCAAGCTTCCCAAGGAATTTGATTTCCGGGAGCAGGAAAGCGCAACTTTAGGC 120
DB 61 TGGATTGGACAGTATTTCTTCAATCAATCTACGCGAGCATGGTTCTGGTAACACTGGA 120
QY 121 GTTACCAATGCAATTCGATATTTGGGTGTAAGAGCTGCTGCTGCTATAGCCGGAGAT 180
DB 121 ACACCAACACCTTCGCGATTTTAGGTAAGAAAGCTGGTATGCAACCTTTTGTGATTGAC 180
QY 181 ATTTTGAAGGACACTGGCAACTGCATTCCTTTTCTCATGCATCT---TGATATTAC 237
DB 181 TTTTTCAAAGGAACCTTAGCAAGCTGCTTCGGATTTATTTTTCATCAAGCGGTTTCT 240
QY 238 CCCTTCTTTCGAGAGCTTTTGGGTTTATAGCCAGCTGTTTCCCATCTTCGCGCAAAATTT 297
DB 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCGCATACCTTCCCTATCTTTCGAGATT 300
QY 298 AAAGCGGTAAAGCGTGGGACATCAGGAGCGGTTTTCGATATTTTACGACCCCTGTTA 357
DB 301 AAAGGTGTAAGGCTGTCGCAACAGTGTGAGAGTATTTTCGGATTTTCGCGCTATCTTC 360
QY 358 TTATACAGATGTTGGGTATTTCTCATCTTTTATATCTGACTAAATTTGTTCTCTC 417
DB 361 TGTCTTACCTTCGGATTAFTCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
QY 418 TCATCATGTTAAAGGATCTATACCTGTTAT 449
DB 421 TCTAGTGTACACCATCGATCGCGGCTGTTAT 452

RESULT 5
US-10-068-080-2
; Sequence 2, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae


```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8039
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; NAME/KEY: CDS
; FEATURE:
; LOCATION: (1)...(609)
US-09-815-242-8039

Query Match      15.6%; Score 90.6; DB 10; Length 609;
Best Local Similarity 48.9%; Pred. No: 6.3e-16;
Matches 298; Conservative 0; Mismatches 284; Indels 27; Gaps 1;

Qy 1 ATGTTAATGCTTTATGATGATTTTGGCCCTACTTGTAGGAGCAATCCATCTGGCTTA 60
Db 1 ATGATGATAATCGTCATCTACTACTAAGTTATCTTATCGGCGCTTTCCCAAGTGGAATC 60
Qy 61 ATTTGGGCAAGCTTGCACCAAGGATTCATATTCGGGAGCACGGAAGCGGCAACTTAGGC 120
Db 61 GTAAATGGAAATATATTTTCAAAAAGATATATAGACAATTTGGTAGTAACTACTGCG 120
Qy 121 GCTACCAATGCAATCCGTCACATTTGGGTGTAAGAGCTGCTCGTCATAGCCGGAGAT 180
Db 121 GCTACTAATAGCTTTAGAGTATAGGTCGCTCGCAGGATTCCTGGTAACATTTCTAGAT 180
Qy 181 ATTTTGAAGGACACTGGCAACTGCATTTGCTTTTCATGCAATGTTGA----- 230
Db 181 ATTTTCAAAAGGTTTCATACCTGTTTCTTCCCTTTATGTTTACCAGTTTCACGCAGATGGC 240
Qy 231 -----TATTCACCCGCTTCTTCAGGAGTCTTTGCGGTTTATAGGCCAC 273
Db 241 CCTATTAGTACTTTTTTACAAATGGTTTAAATTTGGCTTATCTCGCTTACTTGGACAC 300
Qy 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCGGTGGCGACATCAGGAGCGTT 333
Db 301 GTTTATCTGTTTATTTAAATTTCAAGGTGCAAGCGGTGCAACTAGTGCAGGTGTC 360
Qy 334 TTGCTATTTTACGACCCCTCTTATTTATCAGATGGTTGCGGTATTTCTTCACTTTTGA 393
Db 361 GTCTTGGGAGTCAACCCGATCTTTTACTAATATCTTGAATATCTTCTTTATTGTTATG 420
Qy 394 TACTTGACTAAATTTGTTTCTCTCTCATCGATGTTTAAACAGGATCTATCTATATATAT 453
Db 421 AAGATTTTAAATATGTTTCTTTAGCAAGTATCGTTGACGCAATTTGCTGTGATGGC 480
Qy 454 AGTTTCTTTGTCATGATACGATTTATTTAGTTTGTCTTACCCCTGCTCAGTATTTTGTG 513
Db 481 TCGCTTATCATTCAGACTATATTTTATTAGTCGTTAGTCTTCTTCTTCAATCATATTG 540
Qy 514 ATATACACACACCGGAGCAATTTAAACGAATTTATCAATAAACAAGAACCTTAAAGTAAA 573
Db 541 ATAATTAGACATCGCTCTAATATCTCAAGAATTTTATAGAGCGGAAGAACCTTAAATAA 600
Qy 574 TGGTTATAA 582
Db 601 TGGATGTAA 609
```

RESULT 8

```
US-09-815-242-4250
; Sequence 4250, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
```

```
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4250
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4250
```

```
Query Match      15.2%; Score 88.2; DB 10; Length 606;
Best Local Similarity 48.8%; Pred. No. 2.9e-15;
Matches 295; Conservative 0; Mismatches 283; Indels 27; Gaps 1;
```

```
Qy 1 ATGTTAATGCTTTATGATGATTTTGGCCCTACTTGTAGGAGCAATCCATCTGGCTTA 60
Db 1 ATGATGATAATCGTCATCTACTACTAAGTTATCTTATCGGCGCTTTCCCAAGTGGAATC 60
Qy 61 ATTTGGGCAAGCTTGCACCAAGGATTCATATTCGGGAGCACGGAAGCGGCAACTTAGGC 120
Db 61 GTAAATGGAAATATATTTTCAAAAAGATATATAGACAATTTGGTAGTAACTACTGCG 120
Qy 121 GCTACCAATGCAATCCGTCACATTTGGGTGTAAGAGCTGCTCGGTGTCATAGCCGGAGAT 180
Db 121 GCTACTAATAGCTTTAGAGTATAGGTCGCTCGTCAGGATTCCTGGTAACATTTCTAGAT 180
Qy 181 ATTTTGAAGGACACTGGCAACTGCATTTGCTTTTCATGCAATGTTGA----- 230
Db 181 ATTTTCAAAAGGTTTCATACCTGTTTCTTCCCTTTATGTTTACCAGTTTCACGCAGATGGC 240
Qy 231 -----TATTCACCCGCTTCTTCAGGAGTCTTTGCGGTTTATAGGCCAC 273
Db 241 CCTATTAGTACTTTTTTACAAATGGTTTAAATTTGGCTTATCTCGCTTACTTGGACAC 300
Qy 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCGGTGGCGACATCAGGAGCGTT 333
Db 301 GTTTATCTGTTTATTTAAATTTCAAGGTGCAAGCGGTGCAACTAGTGCAGGTGTC 360
Qy 334 TTGCTATTTTACGACCCCTCTTATTTATCAGATGGTTGCGGTATTTCTTCACTTTTGA 393
Db 361 GTCTTGGGAGTCAACCCGATCTTTTACTAATATCTTGAATATCTTCTTTATTGTTATG 420
Qy 394 TACTTGACTAAATTTGTTTCTCTCTCATCGATGTTTAAACAGGATCTATCTATATATAT 453
Db 421 AAGATTTTAAATATGTTTCTTTAGCAAGTATCGTTGACGCAATTTGCTGTGATGGC 480
Qy 454 AGTTTCTTTGTCATGATACGATTTATTTAGTTTGTCTTACCCCTGCTCAGTATTTTGTG 513
Db 481 TCGCTTATCATTCAGACTATATTTTATTAGTCGTTAGTCTTCTTCTTCAATCATATTG 540
Qy 514 ATATACACACACCGGAGCAATTTAAACGAATTTATCAATAAACAAGAACCTTAAAGTAAA 573
Db 541 ATAATTAGACATCGCTCTAATATCTCAAGAATTTTATAGAGCGGAAGAACCTTAAATAA 600
```

Qy	574	TGGTT	578
Db	601	TGCAT	605

```

RESULT 9
US-09-815-242-6900
/ Sequence 6900, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA-OLIA
/ CURRENT APPLICATION NUMBER: US/09/815
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 3
/ SEQ ID NO 6900
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Haemophilus influenzae
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)...(600)
US-09-815-242-6900

```

Db 318 AGGTGTTGCAAC 329

```

RESULT 10
US-09-974-300-4100
; Sequence 4100, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4100
; LENGTH: 72
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-4100

```

RESULT 11
US-09-815-242-7356
; Sequence 7356, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7517
LENGTH: 663
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(663)
US-09-815-242-7517

Query Match 6.8%; Score 39.8; DB 10; Length 663;
Best Local Similarity 55.4%; Pred. No. 0.085;
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 5 TAATGCTTTTATGATATTTTGGCCCTACTTGATAGGCACCATTCCTGCTTAATTG 64
Db 32 TCAATGATGATTTTCCACCTTTTGGGCTATTTGATGGGGGATTCCTTTTGGCTATGCGT 91
OY 65 TGGCGAAGCTTGCACAAAGAAATGATATTCGGGACGACGCAACTTAGGGCGCTA 124
Db 92 TAATGAAATCTTTTATGGCATCGATATTACTAAATCGGATCGGGGGTATTGGCCCGA 151
OY 125 CCAATGCATTCGTCACATT 143
Db 152 CGAATGCTTACGGCTTT 170

RESULT 15
US-09-813-320-3
Sequence 3, Application US/09813320
Patent No. US20020142378A1
GENERAL INFORMATION:
APPLICANT: ZHANG, Hongyu et al.
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001172
CURRENT APPLICATION NUMBER: US/09/813,320
CURRENT FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 397658
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature

LOCATION: (1)...(397658)
OTHER INFORMATION: n = A,T,C or G
US-09-813-320-3

Query Match 6.5%; Score 38; DB 10; Length 397658;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

OY 282 CATCTTTCGCCCAAAATTTAAAGCGGTAAAGCGGTGGGACACATCAGGAGCGCTTTTGGCTATT 341
Db 331668 CTTCTTATGCAATTTTCTAAGTCTCTAACTGGTTGCCCTTATCTATTATGATAGTCTTT 331727
OY 342 TTACGCACCCCTCTTATTTATACGATGGTTGCGGTATTTCTTCATCTTTTATATCTGAC 401
Db 331728 TTACAAAGTTTAAATTTATTAACACATGCAATTTGCACTCTTCTTTCATGTTAAGTCACITGAG 331787
OY 402 TAAATTTGTTCTCTCTCATCGATGTTAAACAGGATCTATCTGTTATATATATAGTTTCTT 461
Db 331788 GAATTTGGGTTTATTAAACTTTAGCTAAATAAACAACAGTCAGTCGCCCTAAGATATATTGTACA 331847
OY 462 TGTCCATGAT 471
Db 331848 AATACCTGAT 331857

Search completed: December 26, 2002, 00:37:24
Job time : 459.89 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 24, 2002, 22:13:55 ; Search time 11247.8 Seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaagtatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	GenEmbl.*	Query				ID	Description
		Result No.	Score	Match	Length		
	1: gb.ba.*	C 1	67.2	10.5	7218	6	I66494
	2: gb.htg.*	C 2	49	7.6	163034	2	AC099432
	3: gb.in.*	C 3	48.2	7.5	171173	2	AF481054
	4: gb.om.*	C 4	45	7.0	85786	9	AL162582
	5: gb.ov.*	C 5	44.6	6.9	115489	2	AC117072
	6: gb.pat.*	C 6	44.4	6.9	179553	2	AC024253
	7: gb.ph.*	C 7	44.4	6.9	186935	2	AC022322
	8: gb.pl.*	C 8	44.4	6.9	176351	2	AC117007
	9: gb.pr.*	C 9	43.8	6.8	161799	2	AC098991
	10: gb.ro.*	C 10	43.8	6.8	211027	2	AC130161
	11: gb.sts.*	C 11	43.5	6.8	62352	2	AC116990
	12: gb.sy.*	C 12	43.5	6.8	201097	2	AC112091
	13: gb.un.*	C 13	43.4	6.8	167439	2	AC102515
	14: gb.vi.*	C 14	43.4	6.8	207364	2	AL671857
	15: em.ba.*	C 15	43	6.7	173048	5	AL772154
	16: em.fun.*	C 16	43	6.7	196628	2	AL772388
	17: em.in.*	C 17	42.4	6.6	161546	2	AC108270
	18: em.mu.*	C 18	42	6.5	3192	3	PF837C3
	19: em.or.*	C 19	42	6.5	3769	3	PF818C1
	20: em.ov.*	C 20	42	6.5	14955	3	AF000580
	21: em.ph.*	C 21	42	6.5	79981	2	AC126315
	22: em.pl.*	C 22	42	6.5	197149	2	AC103297
	23: em.ro.*	C 23	41.5	6.5	321003	2	PF8AL4P3
	24: em.sts.*	C 24	41.4	6.4	198653	2	AC114691
	25: em.sy.*	C 25	41.2	6.4	167364	10	AC121590
	26: em.un.*	C 26	41.2	6.4	220275	10	AL732404
	27: em.vi.*	C 27	40.3	6.4	79769	2	AC115678
	28: em.ba.*	C 28	40.3	6.4	169163	2	AC115666
	29: em.fun.*	C 29	40.5	6.3	56152	2	AC116963
	30: em.in.*	C 30	40.5	6.3	88037	2	PF8AL13P8
	31: em.mu.*	C 31	40.4	6.3	128230	9	HS313L4
	32: em.or.*	C 32	40.4	6.3	131495	2	AC120683
	33: em.ov.*	C 33	40.4	6.3	169965	2	AC118402
	34: em.ph.*	C 34	40.2	6.3	5145	2	AC015315
	35: em.pl.*	C 35	40.2	6.3	43993	2	AC116965
	36: em.ro.*	C 36	40.2	6.3	84785	2	AC117005
	37: em.sts.*	C 37	40.2	6.3	165337	2	AC096212
	38: em.sy.*	C 38	40.2	6.3	182080	3	AC012373
	39: em.un.*	C 39	40.2	6.3	188755	9	AC013562
	40: em.vi.*	C 40	40.2	6.3	303760	3	AE003509
	41: em.ba.*	C 41	39.8	6.2	136431	9	AC093524
	42: em.fun.*	C 42	39.8	6.2	161769	2	AC122097
	43: em.in.*	C 43	39.8	6.2	194020	10	AC121808
	44: em.mu.*	C 44	39.8	6.2	199133	2	AC122212
	45: em.or.*	C 45	39.6	6.2	84563	5	AC096885

ALIGNMENTS

RESULT 1
I66494/c
LOCUS I66494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:272471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 7218)
Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```

source      1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match      10.5% Score 67.2; DB 6; Length 7218;
Best Local Similarity 5.3%; Pred. NO. 2.2e-07;
Matches 21; Conservative 226; Mismatches 149; Indels 0; Gaps 0;

QY 158 CATACCGTGGAAACACTAACTGAAAGTTTCCTGGGATCGTGGCAGCAAGCGCTAAT 217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1476 CCTATCTATGCAATAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRR 1417
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 218 AAAAAGTAGATGTTCCGCAAGAGAGAGATGAGAAACCTGAAACCGCAATAGCGGTAT 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 278 GGAAGGATAGAAACGCTCTAAATTTCCACATTTCCGACAGCGTTGGTCACGACCTCACT 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 AAAGCCCTAAACGGGATAGAGACAGATGGAACCTATAGAGAAACCTCGACAGA 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 TAGAACCGTCTACTAAAGTCACAGATCACAGTGTCTGTAGCTAACGCGCAATAGCCCC 457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 AAGCAGAGAAAGGTGAAACCAAAATAGGACTCATTTGATCTAGAGAGAGTAGCGAT 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 AATAGATCGTGAACGATCAACTAATAGTAGCAG 553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
AC099432/c
LOCUS      AC099432      163034 bp      DNA      linear      HTG 12-JUL-2002
DEFINITION Rattus norvegicus clone CH230-103G20, *** SEQUENCING IN PROGRESS
            *** 82 unordered clones.
ACCESSION  AC099432
VERSION     AC099432.3 GI:21728595
KEYWORDS    HTG; HTGS, PHASE1.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 163034)
AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dedrich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,
            Dearthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loughseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okuwonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
            Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
            Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
            Direct Submission
            Unpublished
            2 (bases 1 to 163034)
            Worley,K.C.
            Direct Submission
            Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 163034)
            Worley,K.C.
            Direct Submission
            Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Jul 11, 2002 this sequence version replaced gi:17973419.

            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GHWE
            Center clone name: CH230-103G20
            ----- Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 79006 bases at least Q40
            Consensus quality: 84121 bases at least Q30
            Consensus quality: 88465 bases at least Q20
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 82 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1382: contig of 1382 bp in length
            * 1383: gap of unknown length
            * 1483: contig of 1134 bp in length
            * 2617: gap of unknown length
            * 2717: contig of 1333 bp in length
            * 4049: contig of 1495 bp in length
            * 4150: gap of unknown length
            * 5644: contig of 1495 bp in length
            * 5744: gap of unknown length
            * 5645: contig of 1169 bp in length
            * 5745: gap of unknown length
            * 6913: contig of 1375 bp in length
            * 7014: gap of unknown length
            * 8388: gap of unknown length
            * 8489: contig of 1650 bp in length

```

```
* 10139 10238: gap of unknown length
* 10239 11734: contig of 1496 bp in length
* 11735 11834: gap of unknown length
* 11835 13287: contig of 1453 bp in length
* 13288 13387: gap of unknown length
* 13388 14709: contig of 1322 bp in length
* 14710 14809: gap of unknown length
* 14810 15861: contig of 1052 bp in length
* 15862 15961: gap of unknown length
* 15962 17518: contig of 1557 bp in length
* 17519 17618: gap of unknown length
* 17620 19300: contig of 1682 bp in length
* 19301 19400: gap of unknown length
* 19401 20729: contig of 1329 bp in length
* 20730 20829: gap of unknown length
* 20830 21939: contig of 1100 bp in length
* 21940 22029: gap of unknown length
* 22030 23415: contig of 1386 bp in length
* 23416 23515: gap of unknown length
* 23516 25034: contig of 1519 bp in length
* 25035 25134: gap of unknown length
* 25135 26213: contig of 1079 bp in length
* 26214 26313: gap of unknown length
* 26314 27705: contig of 1392 bp in length
* 27706 27805: gap of unknown length
* 27806 29246: contig of 1441 bp in length
* 29247 29346: gap of unknown length
* 29347 30523: contig of 1177 bp in length
* 30524 30624: gap of unknown length
* 30624 32166: contig of 1543 bp in length
* 32167 32266: gap of unknown length
* 32267 33674: contig of 1408 bp in length
* 33675 33774: gap of unknown length
* 33775 35023: contig of 1249 bp in length
* 35024 35123: gap of unknown length
* 35124 37010: contig of 1887 bp in length
* 37011 37110: gap of unknown length
* 37111 38878: contig of 1768 bp in length
* 38879 38978: gap of unknown length
* 38979 40329: contig of 1351 bp in length
* 40330 40429: gap of unknown length
* 40430 41896: contig of 1467 bp in length
* 41897 41996: gap of unknown length
* 41997 43475: contig of 1479 bp in length
* 43476 43575: gap of unknown length
* 43576 44684: contig of 1109 bp in length
* 44685 44784: gap of unknown length
* 44785 46637: contig of 1853 bp in length
* 46638 46737: gap of unknown length
* 46738 47987: contig of 1250 bp in length
* 47988 48087: gap of unknown length
* 48088 49155: contig of 1068 bp in length
* 49156 49255: gap of unknown length
* 49256 50589: contig of 1334 bp in length
* 50590 50689: gap of unknown length
* 50690 51988: contig of 1299 bp in length
* 51989 52088: gap of unknown length
* 52089 53966: contig of 1878 bp in length
* 53967 54066: gap of unknown length
* 54067 55332: contig of 1266 bp in length
* 55333 55432: gap of unknown length
* 55433 56979: contig of 1547 bp in length
* 56980 57079: gap of unknown length
* 57080 58450: contig of 1371 bp in length
* 58451 58550: gap of unknown length
* 58551 61272: contig of 2722 bp in length
* 61273 61372: gap of unknown length
* 61373 62639: contig of 1267 bp in length
* 62640 62739: gap of unknown length
* 62740 64180: contig of 1441 bp in length
* 64181 64280: gap of unknown length
* 64281 65566: contig of 1286 bp in length
* 65567 65666: gap of unknown length
```

```
* 65667 66958: contig of 1292 bp in length
* 66959 67058: gap of unknown length
* 67059 68733: contig of 1675 bp in length
* 68734 68833: gap of unknown length
* 68834 70507: contig of 1674 bp in length
* 70508 70607: gap of unknown length
* 70608 71880: contig of 1273 bp in length
* 71881 71980: gap of unknown length
* 71981 73275: contig of 1295 bp in length
* 73276 73375: gap of unknown length
* 73376 74950: contig of 1575 bp in length
* 74951 75051: gap of unknown length
* 75052 76776: contig of 1726 bp in length
* 76777 76876: gap of unknown length
* 76877 77949: contig of 1073 bp in length
* 77950 78049: gap of unknown length
* 78050 79783: contig of 1734 bp in length
* 79784 79883: gap of unknown length
* 79884 82083: contig of 2200 bp in length
* 82084 82183: gap of unknown length
* 82184 83593: contig of 1410 bp in length

Query Match 7.6%; Score 49; DB 2: Length 163034;
Best Local Similarity 44.6%; Pred. No. 0.015;
Matches 193; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

Qy 138 GTAATATCCATTCTTTTCGACCATACCGTTGGAAACACTAACTGAAAGTTCCTTCTGGGA 197
Db 154783 GCAACAAACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACA 154724

Qy 198 TCGTTGGCAGGAGGCTTAATAAAAAAGTAGATGTTCCGCAAGAGGAGAGTAGAACCTGA 257
Db 154723 GCACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 154664

Qy 258 AAACCGACAATAGCCGGTATGGAAGGATAGAAACGCTCTAAATTTCCACCATTTCCGACA 317
Db 154663 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 154604

Qy 318 CGTTGTCAGCAGCTTACTTAAAGCCTAAACGGGATAGAACAGACAGATGGAACCTTA 377
Db 154603 GCAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 154544

Qy 378 ATAGAAGAAACCTCGAGAGATAGAACGCTCATCTACTAAAGTGACAGATGACAGTCTCAG 437
Db 154543 ACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 154484

Qy 438 CTAAAGCCGACAATAGCCCCCAAGCAGAGAAAGGTGAAAAACAAAATAGGACTCATTTGAT 497
Db 154483 GCAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 154424

Qy 498 ACTGAGAGACAAGTAGGCGTAATAGATCTGTGAACGATCAAACTAATAGTAAGCAGTATT 557
Db 154423 GCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 154364

Qy 558 CCTGTTATATCGA 570
Db 154363 ACGTCTTATGGA 154351
```

RESULT 3

```
AF481054 171173 bp DNA linear HTG 06-MAY-2002
LOCUS Mus musculus chromosome 17 clone 201h10 strain 129S6/SVEvtac, ***
DEFINITION SEQUENCING IN PROGRESS ***, 4 ordered pieces.
ACCESSION AF481054
VERSION AF481054.1 GI:20453846
KEYWORDS HTG: HTGS_PHASE2.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 171173)
AUTHORS Brathwaite,M., Weeltz,P., Qian,Y., Dudekula,D., Schlessinger,D. and
```


2276. .2450

2877. .3
71PC-100

3255. 3

4460. .4

compleme

compleme

/rpt_fan

complement
/rpt fan

-25-

```

/rpt_family="SINE/MIR"
11571..11976
/rpt_family="LTR/MaLR"
complement(11977..12042)
/rpt_family="SINE/B4"
12293..12448
/rpt_family="Low_complexity"
complement(14650..14904)
/rpt_family="SINE/B4"
14933..15088
/rpt_family="SINE/B2"
15091..15131
/rpt_family="Simple_repeat"
15235..15323
/rpt_family="Simple_repeat"
15340..15495
/rpt_family="SINE/B2"
15496..15523
/rpt_family="Simple_repeat"
15524..15548
/rpt_family="SINE/B2"
15549..15587
/rpt_family="Simple_repeat"
complement(15647..15804)
/rpt_family="LINE/L1"
16100..16194
/rpt_family="Simple_repeat"
16212..16273
/rpt_family="SINE/B2"
17096..17158
/rpt_family="Simple_repeat"
17158..17210
/rpt_family="Simple_repeat"
17237..17271
/rpt_family="Low_complexity"
complement(17305..17585)
/rpt_family="SINE/B4"
complement(18145..18193)
/rpt_family="DNA/MER1_type"
18201..18235
/rpt_family="Simple_repeat"
complement(18480..18677)

Query Match      7.5%; Score 48.2; DB 2; Length 171173;
Best Local Similarity 48.1%; Pred. No. 0.025;
Matches 168; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

Qy 161 ACCGTTGAAACACTACTACTGAAAGTTTCCTGGGATCGTTGGGACGAAGGCTAATAAA 220
    || || || || || || || || || || || || || || || || || || || || ||
Db 2992 AACAGTGACACACTAACAAGAACTGTGAAAGCAGCAAAAAGTGACACAGCAATAATAAC 3051

Qy 221 AAGTAGATGTTCCGCAAGAGGAGAGTAGAAGCTGAAACCGACATAGCCGGTATGGA 280
    || || || || || || || || || || || || || || || || || || || || ||
Db 3052 AACAGCAACAGCAACATGACAGTAGGACAAACACAGTGTGACACAGCAACAGTAAC 3111

Qy 281 AGGAGTAAACAGCTCTAAATTTCCAGCATTCCTCCAGACGCTGTGTCAGCACTCACTAAA 340
    || || || || || || || || || || || || || || || || || || || || ||
Db 3112 AGTCACAGCAGCATGGCANTACCAACAGCAACAGCAAGTGTGACATGTATACAGTAGTAAC 3171

Qy 341 AGCCTAAACGGGATAGAGACAGAGATGGAACCTTAATAGAGAAACCTCGAGAGATAG 400
    || || || || || || || || || || || || || || || || || || || || ||
Db 3172 AACAGTAACTGCACCACCAACACAGTAAGAAG---AACACACAGCAACAGTGACAGCAGC 3228

Qy 401 AACCGTCATACCTAAAGTCACAGATCACAGTGTCTGTACTACCGCGACATAGCCCAAG 460
    || || || || || || || || || || || || || || || || || || || || ||
Db 3229 TACAGGAGCAGCAATGTAAACAGTAGGACACAGCAACCACTGTGTGACACAGCAGAAAT 3288

Qy 461 ACAGAAAGGTGAAAAACCAAAATAGGACTCATTGTACTAGAGAGAA 509
    || || || || || || || || || || || || || || || || || || || || ||
Db 3289 AACAGTAGACACACAAAGCTACAGCAGTAGTACAGCAACTGTAGAGCA 3337

RESULT 4

```

AL162582
LOCUS
DEFINITION
Human DNA sequence from clone RPI-276N6 on chromosome 6, complete sequence.
ACCESSION
AL162582
VERSION
AL162582.30
KEYWORDS
GI:17907199
SOURCE
HTG.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 85786)
Ellington, A.
Direct Submission
Submitted (16-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
Cloned requests: clonerequest@sanger.ac.uk
On Dec 18, 2001 this sequence version replaced gi:16972822.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/ This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RPI-276N6 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RPI-276N6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RPI-276N6 is at 1 in this sequence. The true left end of clone RPI-366M24 is at 83787 in this sequence.

FEATURES
Location/Qualifiers
1..85786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RPI-276N6"
/clone_lib="RPI-1"
complement(23451..24588)
/note="Sequence from AC024253 sequenced by WUGSC"
24589
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 3800bp by restriction digest data."
34942..34970
/note="Sequence confirmed by AC022322 sequenced by WUGSC."
34971..35889
/note="Sequence from AC022322 sequenced by WUGSC."
35890..36021
/note="Sequence confirmed by AC022322 sequenced by WUGSC."
36022..36030
/note="Sequence from AC022322 sequenced by WUGSC."
36031..36158
/note="Sequence confirmed by AC022322 sequenced by WUGSC."
44819..44856
/note="Sequence from overlapping clone RP11-4J16"


```
/db_xref="GI:20976560"
/translation="MENKROKDEIVLVNGSDTTPNKKQKNELOWSSGGGREGILG
TERTSKLEHPIWOLIGHKEIYSCAFNSYGTALASGGSDKEIFLANVYGEICINYSVLK
GHKGITLHWTSDNEIYACTDKISYGWDSNGKELIKRIEHSYVNSCCPARGP
PLVAGSDRSARIDTRSGSTHLFQHKIPVTSVCFSDASDQLITGGIDNVRKTFNG
RNOEDPLYTLAQHODTITSTSVKDGAYLLSNMDSNCKINDIRPYAPPNNRIKTFNG
AONPEKNIKSSSIDRRICGSSDROVYIWDNTKOLQYCLPGHSGSTNEVTFHP
NEPIIASGSSDKTIVLGEIKP"
CDS
13514..13589
/notes="ORF_ID:dd_02342"
/codon_start=1
/product="Putative GTPase"
/protein_id="AA033150.1"
/db_xref="GI:20976561"
/translation="MEELIKYILVGESGVCKSSILYRFTDNTSQHFAPLGVDFN
VTIRKETGVTKQLMDTAGOERFKSITQFYRSGHGVIVVDTPDKSEKKNW
VEDLIVODTGMIILVGNKSNQMAVKRYTFEGOGEMAEOLKTKLEVSAKENNGVTO
FEDLVODTEATMKNSKVAQNQLNLSVQKRGCC"
CDS
join(15211..15350,15458..16451)
/notes="ORF_ID:dd_01646"
/pseudo
/codon_start=1
complement(join(17012..17403,17496..18273))
/notes="ORF_ID:dd_01647"
/codon_start=1
/product="ROOT IRON TRANSPORTER PROTEIN"
/protein_id="AA033191.1"
/db_xref="GI:20976602"
/translation="MSTLTPPTFLVELCQTAFNNLFKLYTTHDHEEGGGHEGHI
HAGVTEEDPDQYKPIHIAAFVILGVSLFCTWIPILATHIKOLRIPRVAIIYVGS
IGVVLSCAFTHMLPAVLSITSECLPESHEGYEAVPYLFALLAGIMOFIDFVL
OYLTKNEKHLSSNDNISLDVHTPGGEGDHTPGGDEHSKSHGAPGSHVHG
GLMDPAALKTEAYLLERGITVHSVFGLAVGVDDSTKALLVALAFHQFEVGL
GRISDAKTSWHHEALTSIFAVSAPIGIAIGVGVASSLANVGYTLIVGVFDSVC
AGTLYIGSLMIKDPEDMEELCRKYEYFRAGLFLGLVWGAAMAFIKYL"
CDS
join(19682..19960,20123..20209,20353..20607,20740..20846,
20947..20965)
/notes="ORF_ID:dd_01648"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA033151.1"
/db_xref="GI:20976562"
/translation="MKFNKTIISIVIALFTIILQISEITVSNLNNOKDIEFTWYHN
EQDSCKILFKPKLIPHSTNSVDSINFKKAIQIKCNLSLTRESRLYOKSCDLF
KTVVSDKPSNLSSLPKGNLNDSSLOYELKINLDSNNNNIYASVOLKKSPPVT
GDQDOLFIIDFSDILKIVYEEPKVEKEKEKEESQSFYKGYFYLLPLFLII
LVMMAAPPAQOTNAASOGAGSRNNNNN"
CDS
complement(join(21183..21470,21613..21894))
/notes="ORF_ID:dd_01649"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA033152.1"
/db_xref="GI:20976563"
/translation="MNNNPRIKNYKKSLSQSFDDLLVNNKDIIDDSFTCPICLLKLLIDAR
QGCSEGFCCNCTISWLNKNCAVCRHLHSONCLSKNRFLESNIKKIKVPCPNKD
LNDLNDSDGCPLEILEIGNVESHSLCKIHRFYLCQYGGDKGCPVRLNLLSHEIECPF
VELQFNSKNTSLKKKNFRQIPNSPQ"
CDS
complement(22366..23967)
/notes="ORF_ID:dd_01650"
/codon_start=1
/product="Sequence 3 from Patent WO0075321. 6/101"
/protein_id="AA033153.1"
/db_xref="GI:20976564"
/translation="MSAQRFSDLWDFESVVKVVDNGKIFITQOLSFLPSKQOQIES
AYAKSLVKCKDSPAPEYEMGTLRDSFCYQREQLGELIGALHEEFNSRLEKLTIGID
GYLESERQKRLIANGEXTKDLKTAESNESKAKONYEKLARKOEANEDLSKOPPG
AKQKARKTLESATKADGDNEYSVKLCQNOOKFYHEEMPRILDDLORFEVERI
DKSKDMLMEVITQNELVPPAVIHNENIKKGIESTDREDLONYILVYMSGAKPPEA
QVEPYOSGGFPAIVNSSSNLNLSKSGELNGGSIQNGASIISSPOQIQNDHQ
TTPQPNIIQQHQSNNNNNTNNNNVTTPPPPOQOQOLPOTLNNPPQPPIS
LKNSDNINSNNGEIVRALYDYNATEENISFKANALIKVLRIDSESGWQGVIG
ESDRIGVFPNSFISDSSSKRRVDVAGRKCKVLYDRTDCGELNINKEGILTIEYE"
DEGFFGSGNESNVSAARFSPSYVQVI"
complement(join(24934..26525,26681..26777))
/notes="ORF_ID:dd_01651"
/codon_start=1
/product="Sequence 3 from Patent WO0075321. 6/101"
/protein_id="AA033154.1"
/db_xref="GI:20976565"
/translation="MNOVAESINKSLKSTADYTPKGYHITELESDSIYIDEHPSGN
AKIFHPNAVLYEQALAFEGHSATITSGALVTRSGVKTGSPDKRIVKPSODDI
WHPVNIAMDLSLSEMINRERAILDYLNTQEKIYVIGYAGNDPKRIKRVYICARAYHA
LGMNMLLRPANREELRNFPEPDYIYNAGOPPANRYTKGSSSSSIAIDFARKEMVI
LGTQYAGEMKKGIITIMYLMFKMGLPLHSSCNOARNNGDTTLFFGLSGTGKTLISA
DINRELIQDDEHVMTDTCFNIIEGECYAKCIDLSREKEPEIFDAIKFGAVLENVYNE
YSRKVDYNDVSTENTRCAYPLEHINPAKFAIAGHPKNIIMLTCDAFGIILPPVSRLO
ANOVMYHFIQGYTKAVACTEVGTEPTATFSSCYCEPEFVHPHTKYAEMLASOLHKHS
ANOVMLNTGWTTGGSHGVSRILKLAITRAIIDAHSGELEKIPPTMKMDVFGQVPHNSCP
GVPSILPMTGNGWADKKEYIVSTMHLAKLFIENFKKFODKASPELVAAAGPILPO"
CDS
join(28969..29084,29176..29317,29453..30610)
/notes="ORF_ID:dd_00794"
/codon_start=1
Query Match 6.9%; Score 44.6; DB 2; Length 115489;
Best Local Similarity 45.2%; Pred. No. 0.24;
Matches 164; Conservative 0; Mismatches 199; Indels 0; Gaps 0;
QY 215 AATAAAAGCTAGACTGTTCCGCAAGAGGAGACTAGAAACCTCGAAACCCGACAAATAGCCGG 274
DB 89686 AATACCAATAACAGCAATATCAATACCAATAACAAACAGCAATATCAATACACG 89745
QY 275 TATGGAAGGATAGAAACCTCTTAATTTCCACCATTCGACAGCGTTGGTCACGACCTC 334
DB 89746 AATACGACCAACAGCAATACCAACAGCAATATCAATAGCAGCAATACAAACAGCAAT 89805
QY 335 ACTAAAGCCTTAACCGGATAGACAGAGATGGAACGCTAATAGAGAAGAACCTCGAG 394
DB 89806 ACAATCAACAAATAACAAACAGCAATATAAACAACAAATAAACAACAGCAAT 89865
QY 395 AGATAAAGCGTCATACTAAAGTGACAGATCACAGTGTGCTAGTAAACCCGACCAATAGC 454
DB 89866 ATCAATAGCAGCAATAGCAGCAATACCAACAACAACAACAACAACAACAACAACAAC 89925
QY 455 CCCAGACGAGAAAGGTGAAAAACCAAAATAGGACTCATTTGACTGAGAGAGAGTAGC 514
DB 89926 AACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 89985
QY 515 GATAATAGATCGTGAACCATCAAACTAATAGTAGCAGTATTCCTGTTATATCGAGCAT 574
DB 89986 AACATACAAACAACAACAGCAATTCAGCAACAATAATATAACAACAGTAGTAATATCAACAT 90045
QY 575 AGT 577
DB 90046 AGT 90048
RESULT 6
AC024253/c
LOCUS
DEFINITION Homo sapiens chromosome 6 clone rpl1-758C19, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
ACCESSION AC024253
VERSION AC024253.4 GI:9959959
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179553)
AUTHORS Waterston,R.H.
TITLE Direct Submission
```


Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0317M22
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-terminator ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 181940 bases at least Q40
 Consensus quality: 183253 bases at least Q30
 Consensus quality: 183774 bases at least Q20
 Insert size: 186000; agarose-fp
 Insert size: 185435; sum-of-contigs
 Quality coverage: 6.94 in Q20 bases; agarose-fp
 Quality coverage: 7.00 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1363: contig of 1363 bp in length
 * 1364: gap of unknown length
 * 1464 3253: contig of 1790 bp in length
 * 3354 3558: gap of unknown length
 * 4599 4698: contig of 1245 bp in length
 * 4699 5833: gap of unknown length
 * 5834 5933: contig of 1135 bp in length
 * 5934 8646: gap of unknown length
 * 8647 8747: contig of 2713 bp in length
 * 8748 11792: gap of unknown length
 * 11793 11892: contig of 3046 bp in length
 * 11893 14813: gap of unknown length
 * 14814 14913: contig of 2921 bp in length
 * 14914 19837: gap of unknown length
 * 19838 19938: contig of 4924 bp in length
 * 19939 24772: gap of unknown length
 * 24773 24873: contig of 4835 bp in length
 * 24874 30965: gap of unknown length
 * 30966 40841: contig of 6093 bp in length
 * 40842 40941: gap of unknown length
 * 40942 52869: contig of 9776 bp in length
 * 52870 52969: gap of unknown length
 * 52970 63635: contig of 11928 bp in length
 * 63636 63735: gap of unknown length
 * 63736 90222: contig of 10666 bp in length
 * 90223 90323: contig of 26487 bp in length
 * 90324 135921: gap of unknown length
 * 135922 136021: contig of 45599 bp in length
 * 136022 186935: gap of unknown length
 * 186936 186935: contig of 50914 bp in length.

 FEATURES
 source
 1..186935
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-317M22"
 1..1363
 /note="assembly_name:Contig7"
 1464..3253
 /note="assembly_name:Contig8"
 3354..4598
 /note="assembly_name:Contig9"
 4699..5833
 /note="assembly_name:Contig10"
 5934..8646
 /note="assembly_name:Contig11"

misc_feature 8747..11792
 /note="assembly_name:Contig12"
 misc_feature 11893..14813
 /note="assembly_name:Contig13"
 misc_feature 14914..19837
 /note="assembly_name:Contig14"
 misc_feature 19938..24772
 /note="assembly_name:Contig15"
 misc_feature 24873..30965
 /note="assembly_name:Contig16"
 misc_feature 31066..40841
 /note="assembly_name:Contig17"
 misc_feature 40942..52869
 /note="assembly_name:Contig18"
 misc_feature 52970..63635
 /note="assembly_name:Contig19"
 misc_feature 63736..90222
 /note="assembly_name:Contig20"
 clone_end:77
 vector_side:right
 misc_feature 90323..135921
 /note="assembly_name:Contig21"
 misc_feature 136022..186935
 /note="assembly_name:Contig22"
 clone_end:SP6
 vector_side:left
 BASE COUNT 53462 a 40007 c 40377 g 51585 t 1504 others
 ORIGIN
 Query Match 6.9%; Score 44.4; DB 2; Length 186935;
 Best Local Similarity 46.2%; Pred. No. 0.26;
 Matches 147; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
 QY 215 AATAAAGTAGATGTTCCGCAAGAGGAGTAGAAGCTGAAACCGGATAGCGG 274
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 7941 AACACAGCAACAGCAGCAGCAACAATAAGAGCAGCAGCAGCAGCAACAC 8000
 QY 275 TATGGAAGGATAGAAACGCTCTAAATTTCCACATTTCCGACAGCGTTGGT 334
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8001 AGCAGCAACACAGCAGCAACAATAACAAAGCAACGATAGCAACAGTATC 8060
 QY 335 ACTAAAGCCTAAACCGGATAGACAGAGATGGAAGCTTAATAGAGAACTTCG 394
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8061 AGCNTCAGCAACACAGCAGCAACAATAGCAGCAGTAAACAGCAGCAGCAG 8120
 QY 395 AGATAGAACGCTCATCTACTAAAGTGACAGATCACAGTGTGCTAGCTAAC 454
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8121 AGCAGCAGCAGCAGCAACAATAAGAGCAGCAGCAGCAACAGTATAGCAG 8180
 QY 455 CCCAGACGAGAAAGGTGAAAAACCAAAATAGGACTCTTTGATCTAGAGAGAG 514
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8181 AACAGCAGCAACAAACAAAGCAACAATAGCAACAGTATCAGCAACAGCAG 8240
 QY 515 GATATAGAAATCGTGAAC 532
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8241 AACAAAGGATCAGCAAC 8258
 RESULT 8
 AC117007/c
 LOCUS
 Rattus norvegicus clone CH230-358N16, *** SEQUENCING IN PROGRESS
 DEFINITION
 ***, 61 unordered pieces.
 ACCESSION AC117007
 VERSION AC117007.4 GI:21746066
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 176351)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Avele, M., Banks, T.,
 Barbra, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Bunay, C., Burch, P., Burket, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Garrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 176351)
 Worley, K.C.

Direct Submission
 Submitted (06-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 176351)
 Worley, K.C.

Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20136787.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GTFI
 Center clone name: CH230-358N16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 120888 bases at least Q40
 Consensus quality: 127504 bases at least Q30
 Consensus quality: 132322 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 61 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1019: contig of 1019 bp in length
 " 1020 1119: gap of unknown length
 " 1120 2779: contig of 1660 bp in length
 " 2780 2879: gap of unknown length
 " 2880 4496: contig of 1617 bp in length
 " 4497 4596: gap of unknown length
 " 4597 5621: contig of 1025 bp in length
 " 5622 5721: gap of unknown length
 " 5722 6889: contig of 1168 bp in length
 " 6890 8124: contig of 1135 bp in length
 " 8125 8224: gap of unknown length
 " 8225 9804: contig of 1580 bp in length
 " 9805 9904: gap of unknown length
 " 9905 11392: contig of 1488 bp in length
 " 11393 12676: contig of 1184 bp in length
 " 12677 13861: contig of 1085 bp in length
 " 13862 13961: gap of unknown length
 " 13962 15707: contig of 1746 bp in length
 " 15708 15807: gap of unknown length
 " 15808 17582: contig of 1775 bp in length
 " 17583 17682: gap of unknown length
 " 17683 19005: contig of 1323 bp in length
 " 19006 19105: gap of unknown length
 " 19106 21071: contig of 1966 bp in length
 " 21072 22855: contig of 1684 bp in length
 " 22856 22955: gap of unknown length
 " 22956 24423: contig of 1468 bp in length
 " 24424 24523: gap of unknown length
 " 24524 26272: contig of 1749 bp in length
 " 26273 26372: gap of unknown length
 " 26373 28549: contig of 2177 bp in length
 " 28550 30139: contig of 1490 bp in length
 " 30140 30239: gap of unknown length
 " 30240 31817: contig of 1578 bp in length
 " 31818 33762: contig of 1845 bp in length
 " 33763 33862: gap of unknown length
 " 33863 36122: contig of 2260 bp in length
 " 36123 36222: gap of unknown length
 " 36223 38477: contig of 2255 bp in length
 " 38478 38577: gap of unknown length
 " 38578 40590: contig of 2013 bp in length
 " 40591 43154: contig of 2464 bp in length
 " 43155 43254: gap of unknown length
 " 43255 45512: contig of 2258 bp in length
 " 45513 47325: contig of 1713 bp in length
 " 47326 47425: gap of unknown length
 " 47426 49849: contig of 2424 bp in length
 " 49850 49949: gap of unknown length
 " 49950 51626: contig of 1677 bp in length
 " 51627 52857: contig of 1131 bp in length
 " 52858 52957: gap of unknown length
 " 52958 56168: contig of 3211 bp in length
 " 56169 56269: gap of unknown length
 " 56270 59010: contig of 2742 bp in length
 " 59011 61204: contig of 2094 bp in length
 " 61205 61304: gap of unknown length

LOCUS	AC098991	161799 bp	DNA	linear	HTG 12-JUL-2003			
DEFINITION	Rattus norvegicus clone CH230-84N8, *** SEQUENCING IN PROGRESS ***, 71 unordered pieces.							
ACCESSION	AC098991							
VERSION	AC098991.4	GI:21728526						
KEYWORDS	HTG: HTGS_PHASE1.							
SOURCE	Norway rat.							
ORGANISM	Rattus norvegicus							
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.							
	1 (bases 1 to 161799)							
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alzbrooks,S.L., Amaralunga,H.C., Arc,J.R., Ayele,M., Banks,T., Earls,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D., Fouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Euhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.M., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwio,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Stitton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tanssey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.							
	Submitted (08-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA							
REFERENCE	3 (bases 1 to 161799)							
	Worley,K.C.							
AUTHORS	Direct Submission							
	Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA							
COMMENT	Of Jul 11, 2002 this sequence version replaced gi:17973510.							
	***** Genome Center							
Center: Baylor College of Medicine								

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

BASE COUNT

24334 a

7083 c

6689 g

24146 t

100 others

ORIGIN

Query Match

Best Local Similarity

Matches 163; Conservative

0; Mismatches 199; Indels

0; Gaps

0;

0Y

235

CARAGAGGAGTAGAAGCTGAAACCGGACCAATAGCCGCTATCGGAGGGATGACAAACGT

294

Db

13190

CAATAACAATAAATAGTTCCGAAATATAGCAACACAGCAATACAGTAATAGAAATCTC

13131

0Y

295

CCTAAATTTCCACCATTTCCGACAGCGTTGGTCAGCGCTCACTAAACCGCTAAACCGCGA

354

Db

13130

ACCAATTTACACAGTAGTACATGTATAGAAATGACAAAGATCAATTTTATATAGTAA

13071

0Y

355

TAGAAGACAGAGATGGAACGCTTAATAGAAGAAACCTCGAGAGATAGAACCGTCAATAGTAA

414

Db

13070

TAACTCAATAAAGGAATATATATATATATATATATGCAATAATTAATTCAAATTAACAATAATAG

13011

0Y

415

AGTCACAGATCACAGTCTCTAGCTTAACGCGGACCAATAGCCCAAGACGAGAGAGGTGAA

474

Db

13010

ATTCTATGAAAGATAGAGATCAATACATAATAAATCAATATCAATTAATTAATTAATTA

12951

0Y

475

AAACCAAAATAGGACTCATTTGATCTAGCTAGGAGAGAGTAGCGCAATAGATCGTGAACGA

534

Db

12950

TAAATAATAATAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

12891

0Y

535

TCAAACCTAATAGTAAGCAGTATCTCTGTTATATCGACGATAGTTTATTTGACTTTTA

594

Db

12890

TAAATAATAATTAATAATAATAATAATAATAATAATGTTGTTATGTTATCATTTATCAAAATTA

12831

0Y

595

AA 596

Db

12830

TA 12829

RESULT 12

AC112091

LOCUS

DEFINITION

AC112091

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC112091

Rattus norvegicus clone CH230-50E6, *** SEQUENCING IN PROGRESS ***

73 unordered pieces.

AC112091.4 GI:21743615

HTG: HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 201097)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alisbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimige,K., Blankenburg,K., Bonnini,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyie,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homi,F., Howard,S., Huber,J., Hulyk,S., Humel,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,

Lozago,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczka,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 201097)

Worley,K.C.

Direct Submission

Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 201097)

Worley,K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 13, 2002 this sequence version replaced gi:19111309.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc_help@bcm.tmc.edu

----- Project Information

Center project name: GRWY

Center clone name: CH230-50E6

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 157842 bases at least Q40

Consensus quality: 163737 bases at least Q30

Consensus quality: 169262 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 73 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 1004: contig of 1004 bp in length

* 1005 1104: gap of unknown length

* 1105 2146: contig of 1042 bp in length

* 2147 2246: gap of unknown length

* 2247 3645: contig of 1399 bp in length

* 3646 3745: gap of unknown length

* 3746 4955: contig of 1210 bp in length

* 4956 5056: gap of unknown length

* 5056 6381: contig of 1326 bp in length

* 6382 7690: contig of 1209 bp in length

* 7691 7791: gap of unknown length

* 7791 8876: contig of 1086 bp in length

* 8877 8976: gap of unknown length

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 167439)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-pierre, N., Hagos, B., Hartman, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061601.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18947
Center clone name: 425_G_1

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156127 bases at least Q40
Consensus quality: 161465 bases at least Q30
Consensus quality: 163205 bases at least Q20
Insert size: 180000; agarose-ff
Insert size: 164039; sum-of-ctotigs
Quality coverage: 4.2 in Q20 bases; agarose-ff
Quality coverage: 4.6 in Q20 bases; sum-of-ctotigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 699: contig of 699 bp in length
* 700 799: gap of 100 bp
* 800 2116: contig of 1317 bp in length
* 2117 2216: gap of 100 bp

2217 3240: contig of 1024 bp in length
* 3241 3340: gap of 100 bp
* 3341 4407: contig of 1067 bp in length
* 4408 4507: gap of 100 bp
* 4508 5591: contig of 1084 bp in length
* 5592 5691: gap of 100 bp
* 5692 6754: contig of 1063 bp in length
* 6755 6854: gap of 100 bp
* 6855 7975: contig of 1121 bp in length
* 7976 8075: gap of 100 bp
* 8076 9470: contig of 1395 bp in length
* 9471 9570: gap of 100 bp
* 9571 10856: contig of 1286 bp in length
* 10857 10956: gap of 100 bp
* 10957 12838: contig of 1882 bp in length
* 12839 12938: gap of 100 bp
* 12939 14076: contig of 1138 bp in length
* 14077 14176: gap of 100 bp
* 14177 15770: contig of 1594 bp in length
* 15771 15870: gap of 100 bp
* 15871 17857: contig of 1987 bp in length
* 17858 17957: gap of 100 bp
* 17959 20340: contig of 2383 bp in length
* 20341 20440: gap of 100 bp
* 20441 22571: contig of 2131 bp in length
* 22572 22671: gap of 100 bp
* 22672 24605: contig of 1934 bp in length
* 24606 24705: gap of 100 bp
* 24706 27186: contig of 2481 bp in length
* 27187 27286: gap of 100 bp
* 27287 29635: contig of 2349 bp in length
* 29636 29735: gap of 100 bp
* 29736 31986: contig of 2251 bp in length
* 31987 32086: gap of 100 bp
* 32087 34364: contig of 2278 bp in length
* 34365 34464: gap of 100 bp
* 34465 37780: contig of 3316 bp in length
* 37781 37880: gap of 100 bp
* 37881 40408: contig of 2528 bp in length
* 40409 40508: gap of 100 bp
* 40509 44735: contig of 4227 bp in length
* 44736 44835: gap of 100 bp
* 44836 47796: contig of 2961 bp in length
* 47797 47896: gap of 100 bp
* 47897 54045: contig of 6149 bp in length
* 54046 54145: gap of 100 bp
* 54146 60050: contig of 5905 bp in length
* 60051 60150: gap of 100 bp
* 60151 65491: contig of 5341 bp in length
* 65492 65591: gap of 100 bp
* 65592 74597: contig of 9006 bp in length
* 74598 74697: gap of 100 bp
* 74698 82538: contig of 7841 bp in length
* 82539 82638: gap of 100 bp
* 82639 91410: contig of 8772 bp in length
* 91411 91510: gap of 100 bp
* 91511 101578: contig of 10068 bp in length
* 101579 101678: gap of 100 bp
* 101679 111219: contig of 9541 bp in length
* 111220 111319: gap of 100 bp
* 111320 123455: contig of 12136 bp in length
* 123456 123555: gap of 100 bp
* 123556 135268: contig of 11713 bp in length
* 135269 135368: gap of 100 bp
* 135369 167439: contig of 32071 bp in length.

FEATURES

source

1. 167439
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-425G1"
/clone_lib="RPCI-24 Male Mouse BAC"
1. 699
/note="assembly_fragment"

misc_feature

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 150.797 Seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-11

Perfect score: 642

Sequence: 1 tactaatgttatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_l01002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.2	6.3	5688	23 ABL26543	Drosophila melanog
2	40.2	6.3	7588	23 ABL26542	Drosophila melanog
3	38.4	6.0	7337	23 ABL30176	Drosophila melanog
4	38.2	6.0	5613	15 AAO63241	Crucifer 1-aminocy
5	38.2	6.0	7034	22 ABA14938	Human nervous syst
6	38.2	6.0	7036	22 ABA14937	Human nervous syst
7	38	5.9	727	24 ABL29312	Oligonucleotide fo
8	38	5.9	727	24 ABQ29313	Oligonucleotide fo
9	36.8	5.7	5139	21 AAA70139	Plasmodium falcipa

C 10	36.6	5.7	7802	22 AAS45351	Chemically pretrea
C 11	36.6	5.7	7802	24 ABL32765	Human immune syste
C 12	36.6	5.7	7802	24 ABK28184	DNA transcription
C 13	36.4	5.7	807	22 AAL21312	Human breast cance
C 14	36.4	5.7	15428	23 ABL02692	Drosophila melanog
C 15	36.2	5.6	8946	24 ABL32911	Human immune syste
C 16	36	5.6	575	24 ABQ51962	Oligonucleotide fo
C 17	36	5.6	575	24 ABQ51963	Oligonucleotide fo
C 18	36	5.6	15046	24 ABL34101	Human immune syste
C 19	35.8	5.6	5394	24 ABL33008	Human immune syste
C 20	35.6	5.5	640681	24 ABA92787	Buchnera sp. genom
C 21	35.4	5.5	8905	20 AAX20263	Borrelia burgdorfe
C 22	35.2	5.5	534	24 ABQ33814	Oligonucleotide fo
C 23	35.2	5.5	534	24 ABQ33815	Oligonucleotide fo
C 24	34.8	5.4	5935	22 AAS45427	Chemically pretrea
C 25	34.8	5.4	6334	24 ABL33212	Human immune syste
C 26	34.4	5.4	362	22 AAH57352	Human brain specif
C 27	34.4	5.4	803	24 ABQ44790	Oligonucleotide fo
C 28	34.4	5.4	803	24 ABQ44791	Oligonucleotide fo
C 29	34.4	5.4	2054	24 ABK34846	Human cDNA for nov
C 30	34.4	5.4	2306	22 AAK94865	Human full-length
C 31	34.4	5.4	2448	22 AAH57560	Human brain cell s
C 32	34.2	5.3	559	24 ABQ21504	Oligonucleotide fo
C 33	34.2	5.3	559	24 ABQ21505	Oligonucleotide fo
C 34	34.2	5.3	636	23 AAS92306	DNA encoding novel
C 35	34.2	5.3	646	24 ABQ46952	Oligonucleotide fo
C 36	34.2	5.3	646	24 ABQ46953	Oligonucleotide fo
C 37	34.2	5.3	744	23 AAS77979	DNA encoding novel
C 38	34.2	5.3	4590	22 AAH24065	Yeast AOD9604-asso
C 39	34	5.3	694	24 ABQ19008	Oligonucleotide fo
C 40	34	5.3	694	24 ABQ19009	Oligonucleotide fo
C 41	33.8	5.3	9774	22 ABA19485	Human nervous syst
C 42	33.6	5.2	521	24 ABQ37468	Oligonucleotide fo
C 43	33.6	5.2	521	24 ABQ37469	Oligonucleotide fo
C 44	33.6	5.2	664	22 AAK63846	Human immune/haema
C 45	33.6	5.2	5059	20 AAX84332	Stealth virus nucl

ALIGNMENTS

RESULT 1

ABL26543
ID ABL26543 standard; DNA; 5688 BP.

XX ABL26543;

AC ABL26543;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 31102.

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31102.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

KW Drosophila melanogaster.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell


```

/label= Exon 2.
1910..2161
/*tag= j
/label= Intron 2.
2162..2322
/*tag= k
/label= Exon 3.
2323..2595
/*tag= l
/label= Intron 3.
2596..3621
/*tag= m
/label= Exon 4
3926..2931
/*tag= n
/*note= "Putative polypeptide signal."

```

```

/*tag= "III"
/label= Exon 4
3926..2931
/*tag= "n"
/note= "Putative polyadenylation signal."

```

```

FT      poly6_signal
FT      3220.12931
FT      /*tag= n
FT      /note= "putative polyadenylation signal."

```

PN WO9406449-A.
XX
PD 28-APR-1994.
XX
PF 13-OCT-1993: 93WO-US09816.

PD 28-APR-1994.
XX
PF 13-OCT-1993: 93WO-US09816.

PF 13-OCT-1993: 93WO-US09816.

PR 15-OCT-1992; 92US-0962481.

PA (GEHO) GEN HOSPITAL CORP.;

PI Goodman H. Van der Straeten D. Van Montague M:

DK WP1, 1994-150616/18.
DR P-PSDB: AAR53114.

XX	DNA encoding crucifer ACC synthase - used for prodn. of
PT	recombinant polypeptide(s) or regulating ethylene-inducible
PT	events in a plant.

PT recombinant polypeptide(s) or regulating ethylene-inducible events in a plant.

PS Claim 1; Figure 1A; 54pp; English.

The crucifer 1-aminocyclopropane-1-carboxylate (ACC) synthase coding sequence can be used to produce transgenic plants in which ethylene inducible events such as fruit ripening, fruit maturation or senescence are inhibited

sequence can be used to produce transgenic plants in which ethylene inducible events such as fruit ripening, fruit maturation or senescence are inhibited

CC senescence are inhibited.

XX

Sequence	5613 BP:	179A A:	959 C:	960 C:	1900 T:	0 other.
----------	----------	---------	--------	--------	---------	----------

2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 2680, 2681, 26

Query Match 6.0%; Score 38.2; DB 15; Length 5613;
Best Local Similarity 48.8%; Pred. No. 0.63;

Matches 103; Conservative 0; Mismatches 108; Indels 0; Gaps

0v 375 CTTATGACACACACCTCCACACATCATCACCTCTCC 13A

43
 42
 41
 40
 39
 38
 37
 36
 35
 34
 33
 32
 31
 30
 29
 28
 27
 26
 25
 24
 23
 22
 21
 20
 19
 18
 17
 16
 15
 14
 13
 12
 11
 10
 9
 8
 7
 6
 5
 4
 3
 2
 1

Db 434 CTACAAAAAATCAAGATATAAACTAACTCTCACTAGTCACTAATAATTTTC 493

[illegible][illegible]

Db 494 ATTACAATTCATATATGATCTACTACTAACAAATTGTTGGTTATACAAACAAAATTTATT 553

[illegible][illegible]

Db 554 TTTCGTAAGACGGAAATTTTGAAATCAAATTCCTCAACACTCAAATGAATATTCAGTAGT 613

[illegible]

COG TTTT TTCTTCCTCCCTCCTT CCG fA

Db 614 AGTATATCAATGACTAGATTAGATATTTCTT 644

RESULT 5

ABA14938/c

ID ABA14938 standard; DNA; 7034 BP.

AC ARA74938:
XX

[illegible]

DT	23-JAN-2002	(first entry)	PR	14-SEP-2000;	2000US-0232398.
XX			PR	14-SEP-2000;	2000US-0232399.
DE		Human nervous system related polynucleotide SEQ ID NO 7269.	PR	14-SEP-2000;	2000US-0232400.
XX			PR	14-SEP-2000;	2000US-0232401.
KW		Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;	PR	14-SEP-2000;	2000US-0233063.
KW		immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;	PR	14-SEP-2000;	2000US-0233064.
KW		antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;	PR	14-SEP-2000;	2000US-0233065.
KW		antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;	PR	21-SEP-2000;	2000US-0234223.
KW		antiallergic; antidiabetic; anticonvulsant; antifungal;	PR	21-SEP-2000;	2000US-0234274.
KW		antiparasitic; cardiac; immune disorder; cardiovascular disorder;	PR	25-SEP-2000;	2000US-0234997.
KW		neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.	PR	25-SEP-2000;	2000US-0234998.
XX			PR	26-SEP-2000;	2000US-0235484.
OS		Homo sapiens.	PR	27-SEP-2000;	2000US-0235834.
XX			PR	27-SEP-2000;	2000US-0235836.
PN		WO200159063-A2.	PR	29-SEP-2000;	2000US-0236327.
XX			PR	29-SEP-2000;	2000US-0236367.
XX			PR	29-SEP-2000;	2000US-0236368.
XX			PR	29-SEP-2000;	2000US-0236369.
XX			PR	29-SEP-2000;	2000US-0236370.
XX			PR	02-OCT-2000;	2000US-0236802.
XX			PR	02-OCT-2000;	2000US-0237037.
XX			PR	02-OCT-2000;	2000US-0237038.
XX			PR	02-OCT-2000;	2000US-0237039.
XX			PR	02-OCT-2000;	2000US-0237040.
PR			PR	13-OCT-2000;	2000US-0239935.
PR			PR	13-OCT-2000;	2000US-0239937.
PR			PR	20-OCT-2000;	2000US-0240960.
PR			PR	20-OCT-2000;	2000US-0241785.
PR			PR	20-OCT-2000;	2000US-0241786.
PR			PR	20-OCT-2000;	2000US-0241808.
PR			PR	20-OCT-2000;	2000US-0241809.
PR			PR	20-OCT-2000;	2000US-0241826.
PR			PR	20-OCT-2000;	2000US-0242221.
PR			PR	01-NOV-2000;	2000US-0244617.
PR			PR	08-NOV-2000;	2000US-0246474.
PR			PR	08-NOV-2000;	2000US-0246475.
PR			PR	08-NOV-2000;	2000US-0246476.
PR			PR	08-NOV-2000;	2000US-0246477.
PR			PR	08-NOV-2000;	2000US-0246478.
PR			PR	08-NOV-2000;	2000US-0246523.
PR			PR	08-NOV-2000;	2000US-0246524.
PR			PR	08-NOV-2000;	2000US-0246525.
PR			PR	08-NOV-2000;	2000US-0246526.
PR			PR	08-NOV-2000;	2000US-0246527.
PR			PR	08-NOV-2000;	2000US-0246528.
PR			PR	08-NOV-2000;	2000US-0246532.
PR			PR	08-NOV-2000;	2000US-0246609.
PR			PR	08-NOV-2000;	2000US-0246610.
PR			PR	08-NOV-2000;	2000US-0246611.
PR			PR	08-NOV-2000;	2000US-0246613.
PR			PR	17-NOV-2000;	2000US-0249207.
PR			PR	17-NOV-2000;	2000US-0249208.
PR			PR	17-NOV-2000;	2000US-0249209.
PR			PR	17-NOV-2000;	2000US-0249210.
PR			PR	17-NOV-2000;	2000US-0249211.
PR			PR	17-NOV-2000;	2000US-0249212.
PR			PR	17-NOV-2000;	2000US-0249213.
PR			PR	17-NOV-2000;	2000US-0249214.
PR			PR	17-NOV-2000;	2000US-0249215.
PR			PR	17-NOV-2000;	2000US-0249216.
PR			PR	17-NOV-2000;	2000US-0249217.
PR			PR	17-NOV-2000;	2000US-0249218.
PR			PR	17-NOV-2000;	2000US-0249244.
PR			PR	17-NOV-2000;	2000US-0249245.
PR			PR	17-NOV-2000;	2000US-0249264.
PR			PR	17-NOV-2000;	2000US-0249265.
PR			PR	17-NOV-2000;	2000US-0249297.
PR			PR	17-NOV-2000;	2000US-0249299.
PR			PR	17-NOV-2000;	2000US-0249300.
PR			PR	01-DEC-2000;	2000US-0250391.
PR			PR	01-DEC-2000;	2000US-0251160.
PR			PR	05-DEC-2000;	2000US-0251030.


```
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249268.
PR 17-NOV-2000; 2000US-0249269.
PR 17-NOV-2000; 2000US-0249270.
PR 17-NOV-2000; 2000US-0249271.
PR 17-NOV-2000; 2000US-0249272.
PR 17-NOV-2000; 2000US-0249273.
PR 17-NOV-2000; 2000US-0249274.
PR 17-NOV-2000; 2000US-0249275.
PR 17-NOV-2000; 2000US-0249276.
PR 17-NOV-2000; 2000US-0249277.
PR 17-NOV-2000; 2000US-0249278.
PR 17-NOV-2000; 2000US-0249279.
PR 17-NOV-2000; 2000US-0249280.
PR 17-NOV-2000; 2000US-0249281.
PR 17-NOV-2000; 2000US-0249282.
PR 17-NOV-2000; 2000US-0249283.
PR 17-NOV-2000; 2000US-0249284.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249286.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249288.
PR 17-NOV-2000; 2000US-0249289.
PR 17-NOV-2000; 2000US-0249290.
PR 17-NOV-2000; 2000US-0249291.
PR 17-NOV-2000; 2000US-0249292.
PR 17-NOV-2000; 2000US-0249293.
PR 17-NOV-2000; 2000US-0249294.
PR 17-NOV-2000; 2000US-0249295.
PR 17-NOV-2000; 2000US-0249296.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249298.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0250392.
PR 05-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 20C1-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases.
XX
XX Disclosure; SEQ ID NO 7268; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
XX (ABAI1678-ABAI18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (antagonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7036 BP; 2149 A; 1291 C; 1364 G; 2232 T; 0 other;

Query Match 6.0%; Score 38.2; DB 22; Length 7036;
Best Local Similarity 44.2%; Pred. No. 0.68;
Matches 157; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 140 AAANTCCATTCTTTTCGACCATACCGTTGGAAACACTAACTGAAAGTTTCCTTGGATC 199
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4015 AAANTGATAATTATAGAAATTTTATATGGAAAAAGAGTTTAAACAGACAGAGATAGA 3956
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 200 GTTGGCAGGAAGCTAATAAAAGTAGATGTTCCGCAAGAGAGAGAGTAGAAGACCTGAAA 259
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 3955 TTANCAAGGTCAACATAGCTTAAAGGGTCTGGAAGAGAAATTTGAGAGGTGGGA 3896
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 260 ACCGACANTAGCCGGTATGGAAGGATAGAAAGCTCTTAANTTCCACCATCCGACAGC 319
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 3895 AGATGAANTAGACAAAAAGATGATGAAAAAATTTCCAGAAATTTGTAAGATGTAAGTTG 3836
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 320 GTTGGTCACGACCTCACTAAAGCCCTAAACGCGGATAGAGACAGACAGATGGAACGTAAT 379
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 3835 TTAGATCAATCATCAACAAAGTTTGAAGTTAATAAGGAAGAAAGACATATCAGGAAC 3776
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 380 AGAAGAAACCTCGAGAGATAGAACCGCTCTACTTAAAGTAGACAGATCACAGTTCGTAGCT 439
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 3775 TTACTAATCAAGGAAGATGTTAGGACAAATATTAAATATATCAGAAAAAATAGAGTTTCAGGT 3716
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```


PD 06-DEC-2001.
 XX 06-APR-2001; 2001WO-EP03973.
 PF 06-APR-2000; 2000DE-1019058.
 XX 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-090046/12;
 DR
 XX
 XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
 PT tumours or cancer
 XX
 XX Claim 1; SEQ ID NO 58; 32pp; English.
 XX
 XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK2127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 XX Sequence 7802 BP; 1718 A; 478 C; 2409 G; 3197 T; 0 other;
 SQ
 Query Match 5.7%; Score 36.6; DB 24; Length 7802;
 Best Local Similarity 49.2%; Pred. No. 2.1;
 Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 221 AAGTAGATGTTCCGCAAGAGGAGTAGAAGACCTGAAACCCACACATACCGGTATGGA 280
 DB 4669 AAATACCCATTAGCAAAACGTAACCTTAAACCCGCAACCCGCAAAATATCCCGTAA 4610
 QY 281 AGGATAGAACGTCCTTAAATTCACCATTCGACAGCGTTGCTCAGCACCTCACTAAA 340
 DB 4609 ATCCCGACCAAAACCAATTAATTCGATACAAACCCGCAAACTAATTCCTTAATAT 4550
 QY 341 AGCTTAACCCGGATAGACAGAGATGGAACGCTTAATAGAGAACCTTCGAGAGATAG 400
 DB 4549 AAAAAAACCCGCAAAACCAATTAATTAATTCGACAGCAACGCAAAATTAATTA 4490
 QY 401 AACCGTCATATACTAAA 415
 DB 4489 ATCCCTAATAATAA 4475
 RESULT 13
 AAL21312/c
 ID AAL21312 standard; cDNA; 807 BP.
 XX
 XX AAL21312;
 AC

XX 07-DEC-2001 (first entry)
 DT
 XX Human breast cancer expressed polynucleotide 13769.
 DE
 XX Human; breast cancer; cell marker; cytostatic; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200151628-A2.
 PN
 XX 19-JUL-2001.
 PD
 XX 10-JAN-2001; 2001WO-US00798.
 PF
 XX 14-JAN-2000; 2000US-0176077.
 PR 14-MAR-2000; 2000US-0189167.
 PR 24-MAR-2000; 2000US-0192099.
 PR 29-MAR-2000; 2000US-0193480.
 PR 15-MAY-2000; 2000US-0205230.
 PR 09-JUN-2000; 2000US-0211315.
 PR 23-JUL-2000; 2000US-0220534.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Lillie J, Xu Y, Wang Y, Steinmann K;
 PI WPI; 2001-451856/48.
 XX
 XX New peptide useful as a marker for the diagnosis of breast cancer
 PT
 XX Claim 1; Page 2449; 3695pp; English.
 PS
 XX The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 XX
 XX Sequence 807 BP; 90 A; 163 C; 125 G; 418 T; 11 other;
 SQ
 Query Match 5.7%; Score 36.4; DB 22; Length 807;
 Best Local Similarity 44.0%; Pred. No. 1;
 Matches 133; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
 QY 240 AGGAGAGTAGAACCTTGAACCCGCAACATAGCCGGTATGGAAGGGATAGAACGCTCTAA 299
 DB 743 AGAAAG 684
 QY 300 ATTTCACCATTCCTCCGACAGCGTTGGTTCACGACCTCACTAAACGCTTAAACGCGATAGAA 359
 DB 683 AATCAAG 624
 QY 360 GACAGAGATGGAACGCTTAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
 DB 623 GAG 564
 QY 420 CAGATCAGAGTGTGCTAGCTTAACCGCCGCAATAGCCCAAGAGAGAGAGAGAGAGAGAG 479
 DB 563 AAG 504
 QY 480 AAATAGGAGACTCATTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
 DB 503 AAAAAAG 504
 QY 540 CTAATAGTAGCAGTA 555
 DB 443 AAAAAAG 428

RESULT 14
ABL02692/C
ID ABL02692 standard; cDNA; 15428 BP.
XX
AC ABL02692;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2558.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB58589.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 2558; 21pp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15428 BP; 4834 A; 2815 C; 2869 G; 4910 T; 0 other;

Query Match 5.7%; Score 36.4; DB 23; Length 15428;
Best Local Similarity 56.8%; Pred. No. 3.2;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 383 AGAAACCTCAGAGATAGAACCGCTACTACTAAAGTGACAGATCACAGTGTCTAGCTAAC 442
DB 14493 AGNACCATTTAAGCTAGAGGTTGAGATGAGCATACAGATTTCTAGAGACCCCTAAC 14434
QY 443 GCCGACAAATAGCCCCAGAGAGAGAGGTGAAAAACCAAAATAGGACTCATTTGATACT 500
DB 14433 GCCACAAAAACCCCCCAAACTGCAAGGACGAGCTTTTAAACAACTTTTGGATATT 14376

RESULT 15
ABL32911/C
ID ABL32911 standard; DNA; 8946 BP.
XX
AC ABL32911;
XX
DT 26-MAR-2002 (first entry)
XX

DE Human immune system associated gene SEQ ID NO: 884.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 884; 32pp + Sequence Listing: German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 8946 BP; 2493 A; 66 C; 1749 G; 4638 T; 0 other;

Query Match 5.6%; Score 36.2; DB 24; Length 8946;
Best Local Similarity 45.8%; Pred. No. 3;
Matches 123; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 296 CTAAATTTCCACCATTCGACAGCGTTGGTCACGACCTCCTAAAGCCCTAAACGCGGAT 355
DB 2718 CTACACTCCAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 2659
QY 356 AGAAGACAGAGATGGAACGCTAATAGAGAAACCTCGAGAGATAGAACCGTCTACTATAA 415
DB 2658 AAA 2599
QY 416 GTGACAGATCACAGTGTCTAGCTACGCGGACAAATAGCCCCAGAGAGAGAGGTGAAA 475
DB 2598 AAA 2539
QY 476 AACCAAAATAGGACTCATTGATCTAGAGAGAGAGTACGATATAGATTCGGAACGAT 535
DB 2538 AAA 2479
QY 536 CAAACTAATAGTAGCAGTATTCTCTTTATTC 568
DB 2478 AAAAAAAAAAATTAATTATTATTAATTCATTAATCTC 2446

Search completed: December 24, 2002, 22:49:18
Job time : 192.797 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:16:24 ; Search time 3529.69 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgtatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estim:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_othr:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	50	7.8	1621	BE868376	BE868376 601444337
c 2	48.4	7.5	997	CNS0005TE	AL060767 Drosophill
3	46	7.2	1007	CNS006X9S	AL19462 T3 end of
c 4	44.6	6.9	555	AQ577769	AQ577769 nbxb0091L
5	44.2	6.9	589	CNS03012	AL253412 Tetraodon
6	44.2	6.9	789	AG061290	AG061290 Pan trogl

7	44	6.9	494	13	BM641696
8	44	6.9	1077	12	BG495797
9	43.2	6.7	872	17	CNS002JTV
10	43	6.7	827	17	CNS02156
11	42.6	6.6	1689	12	BE872484
12	42.6	6.6	1715	12	BF243003
c 13	42.4	6.6	751	17	AG077983
14	42.2	6.6	679	17	AZ650047
15	42	6.5	712	17	AG056512
16	41.8	6.5	1101	17	CNS0182H
c 17	41.6	6.5	391	14	H43083
18	41.4	6.4	485	12	BF347872
c 19	41.4	6.4	878	9	AL536095
20	41.2	6.4	636	13	BJ335989
c 21	41.2	6.4	858	17	CNS012GG
c 22	40.8	6.4	954	17	CNS0050Y
23	40.8	6.4	1034	12	BF346237
c 24	40.4	6.3	902	17	CNS0060P
25	40.4	6.3	946	17	CNS007YY
26	40.2	6.3	894	17	CNS018BG
c 27	40.2	6.3	983	13	BM415036
c 28	40	6.2	1101	17	CNS0183G
29	39.8	6.2	386	17	BH863527
30	39.8	6.2	1892	12	BG121866
31	39.6	6.2	728	17	BH728612
32	39.4	6.1	494	13	BJ360881
33	39.4	6.1	654	13	BJ328915
34	39.2	6.1	562	17	PT013K11R
35	39.2	6.1	812	17	AG031898
36	39.2	6.1	904	17	AG056296
c 37	39.2	6.1	933	17	AZ166708
38	39.2	6.1	1019	13	BI648486
c 39	39.2	6.1	1101	17	CNS0178P
40	39	6.1	972	17	AG045001
c 41	39	6.1	1101	17	CNS0039G
42	38.8	6.0	366	9	AU262852
c 43	38.8	6.0	660	17	AQ873871
44	38.6	6.0	428	13	BJ391809
c 45	38.6	6.0	910	17	CNS0060N

ALIGNMENTS

RESULT 1
BE868376
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BE868376
601444337F1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3848529 5', mRNA linear EST 20-OCT-2000
mRNA sequence.
BE868376
BE868376.1 GI:10317152
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1621)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LIAW9564 row: i column: 10
High quality sequence stop: 33.
Location/Qualifiers
1. .1621

FEATURES
source

[illegible]

Qy	268	TAGCCGGTATGGAAGGATTAGAAACGTCTTAATTTCCACCATTCCGCACAGCGTTGTCA	327
Db	848	RRRARGAGRGRRGRGRRRRARRRARRAGAARARRRRRARARRRRRARRRRRRRA	789
Qy	328	CGACTTCACCTAAAAGCCCTAAACGCCGATAGAAGACACAGATGGAACCTTAATAACAAGAA	387
Db	788	RGARRRRRGRRRRGRRRRRRARRARARARAGAAAAARRRRARRAGAAAAARRR	729
Qy	388	CCTCGACAGATAGAACCGTCTACTAAGTGACAGATCAGAGTCTCGTAGCTACGCCGA	447
Db	728	RRRRRRRGAGARRRRARRRARRAGAAARRRRRRRGAGARRRRGRGRRRRGAGA	669
Qy	448	CAATAGCCCCAAGACGAGAAGGTGAAAAACCAAATAGGACTCATTTGACTAGAGAG	507
Db	668	RRRRRGRRMTTRARRRRRRRAGAAARRRARGARRRRRRRRRRRRRRRRRRRAGR	609
Qy	508	AAGTAGCGCATTAATAGAAATCGTGAACGATCAAACTAATAGTAAGCAGTA	555

RESULT 3	CNS06X9S	1007 bp	DNA	linear	GSS 06-JUL-2001
LOCUS					
DEFINITION	T3 end of clone AX0AA039F08 of library AX0AA from strain CBS 7064 of <i>Pichia farinosa</i> , genomic survey sequence.				
ACCESSION	AL419462				
VERSION	AL419462.1	GI:12202640			
KEYWORDS	GSS.				
SOURCE	<i>Pichia farinosa</i> .				
ORGANISM	<i>Pichia farinosa</i> .				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; <i>Pichia</i> .				
AUTHORS	1 (bases 1 to 1007) Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bilotin-Fukuhara,M., Bon,E., Bottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Winkler,P. and Weissenbach,J.				

TITLE	Wincker,P. and Weissenbach,J. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL	FEMS Lett. 487 (1), 3-12 (2000)
MEDLINE	20584711
PUBMED	11152876
REFERENCE	2 (bases 1 to 1007)
AUTHORS	de Montigny,J., Spehner,C., Souciet,J., Tekaia,F., Gujon,B., Wincker,P., Artiguenave,F. and Potier,S.
TITLE	Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila
JOURNAL	FEMS Lett. 487 (1), 87-90 (2000)
MEDLINE	20584725
PUBMED	11152890
REFERENCE	3 (bases 1 to 1007)
AUTHORS	Genoscope.

AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (08-SEP-2000) Genoscope -- Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
sagref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT
This GSS is part of a random genomic sequencing program of thirteen
yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotoleros*, *Kluyveromyces*

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and varrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

Location/Qualifiers
1..1007
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX00A039F08"
/clone_lib="AX00A"
/note="end : T3"
533 a 86 c 126 g 161 t 101 others

BASE COUNT

Query Match 7.2%; Score 46; DB 17; Length 1007;
Best Local Similarity 36.0%; Pred. No. 0.013;
Matches 121; Conservative 45; Mismatches 170; Indels 0; Gaps 0;

Qy 214 TAATAAAAGTAGATGCTCCGCAAGAGAGAGTAGAACCTGAAACCGACAATAGCCG 273

Db 164 WMAAANAATAAAGTTTAAANAAGAAAGAGAAATGAAATTAAMWMAAACA 223

Qy 274 GTATGAAGGGATAGAAAGCTCTTAATTTCCACCATTCGGACAGGTTGGTCAGCACCT 333

Db 224 MGTTCATGTCWGGTGATGAWGTTRATWTAWGAWAARAWARAAAGGWTAAAGMA 283

Qy 334 CACTAAAGCCCTAAGCCGGATAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393

Db 284 WAAWAAWAAWAGAGAGGGTATTAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 343

Qy 394 GAGATAGAACCGCTACTACTACTACAGATCACAGTGTCTAGCTTAACCGCCACAATAG 453

Db 344 AAATAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 403

Qy 454 CCCAAGACGAGAAAGGTGAAACCAATAGGACTCTATTGATCTAGAGAGAGAGAGAG 513

Db 404 ATAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 463

Qy 514 CGTATAGATCGTGAACGATCAACTAATAGTAA 549

Db 464 WAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 499

RESULT 4

AQ577769/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AQ577769 555 bp DNA linear GSS 02-JUN-1999
nbx0091L07f CUGI Rice BAC Library Oryza sativa genomic clone
nbx0091L07f, DNA sequence.

ACCESSION AQ577769
VERSION AQ577769.1 GI:4978254
KEYWORDS GSS.

SOURCE Oryza sativa.
ORGANISM Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 555)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCCTATAGG
Class: BAC ends
High quality sequence stop: 101.

FEATURES

source

Location/Qualifiers
1..555
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbx0091L07f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="vector: pBelobAC11; Site 1: HindIII; Site 2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening." 4 others

BASE COUNT 89 a 120 c 86 g 256 t

Query Match 6.9%; Score 44.6; DB 17; Length 555;
Best Local Similarity 54.6%; Pred. No. 0.033;
Matches 89; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 383 AGAAACCTCGAGAGATAGAACCGCTCATCTAAAGTGACAGATCACAGTCTCGCTAC 442

Db 481 AGCAAAAAGAGAGATGGAACATGGATGATAAAACACACAGCTCAAAAAAGATAGA 422

Qy 443 GCGACATAGCCCAAGACGAGAGAGGTGAAAAACCAAAATAGGACTCATTTGATCTCA 502

Db 421 GTAGGACAAAATAGAACGACAGAAAATAAATAAATAAATAAATAAATAAATAA 362

Qy 503 GAGAGAAAGTAGCGTAATAAGAAATCGTGAACGATCAAACTAATA 545

Db 361 AGCAAAAGTAAAGAAAAAGAAATATGAGGAGATATATAAAA 319

RESULT 5
CNS0301Z 589 bp DNA linear GSS 17-MAY-2000

LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone

DEFINITION 043121 of library G from tetraodon nigroviridis, genomic survey

ACCESSION AL253412

VERSION AL253412.1 GI:7974424

KEYWORDS GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

1 (bases 1 to 589)
Ruest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.

TITLE Human gene number estimate provided by genome wide analysis using

JOURNAL Tetraodon nigroviridis DNA sequence

REFERENCE Unpublished

2 (bases 1 to 589)
Ruest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,

(where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 14/15 clones contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 6.6%; Score 42.6; DB 12; Length 1715;
Best Local Similarity 53.6%; Pred. No. 0.14; 1;
Matches 113; Conservative 0; Mismatches 94; Indels 4; Gaps 1;

QY 338 AAAAGCCTAAACGCCGATAGAAAGACAGAGATGGGAACGCTAATAGAGAAACCTCGAGAGA 397

[illegible][illegible]

AG077983/C	AG077983	751 bp	DNA	linear	GSS 03-NOV-2001
LOCUS	Pan troglodytes	DNA, clone: PTB-072k12.F,	genomic survey sequence.		
DEFINITION	AG077983				
ACCESSION	AG077983.1	GI:16629785			
VERSION	GSS.				
KEYWORDS	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male				
SOURCE					

ORGANISM

BAC Library clone: PTB-072K12.F.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

REFERENCE

AUTHORS

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE
BAC end sequences of Library PTB
Unpublished
REFERENCE
2 (bases 1 to 751)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
AUTHORS
Direct Submission
TITLE
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC):
1-7-22 Seihoro-chou, Tsukuba, Ibaraki 305-0856, Japan
(E-mail: chimpbes@gsc.riken.go.jp, URL: <http://hgsc.riken.go.jp/>,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:25:54 ; Search time 93.625 Seconds
(without alignments)
2102.926 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgttatcaaaataa.....tagtctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327.8	51.1	344	US-09-222-938A-15	Sequence 15, Appl
2	100	15.6	582	US-09-222-938A-72	Sequence 72, Appl
3	67.2	10.5	7218	US-08-232-463-14	Sequence 14, Appl
4	38.2	6.0	5613	US-08-463-418-1	Sequence 1, Appl
5	33	5.1	597	US-08-332-766A-19	Sequence 19, Appl
6	33	5.1	1488	US-08-875-972-3	Sequence 3, Appl
7	33	5.1	2275	US-08-743-637B-2	Sequence 2, Appl
8	33	5.1	2275	US-08-526-840B-2	Sequence 2, Appl
9	33	5.1	2764	US-08-923-454A-9	Sequence 9, Appl
10	33	5.1	2764	US-08-832-867-2	Sequence 2, Appl
11	33	5.1	2765	US-08-888-077A-1	Sequence 1, Appl
12	33	5.1	2765	US-08-706-344C-1	Sequence 1, Appl
13	33	5.1	2765	US-08-706-344C-3	Sequence 3, Appl
14	33	5.1	2765	US-08-706-344C-27	Sequence 27, Appl
15	33	5.1	2765	US-08-706-344C-29	Sequence 29, Appl
16	33	5.1	2765	US-08-706-344C-31	Sequence 31, Appl
17	33	5.1	2791	US-08-967-101-1	Sequence 1, Appl
18	33	5.1	2791	US-08-967-101-133	Sequence 133, Appl
19	33	5.1	2791	US-08-592-541-1	Sequence 1, Appl
20	33	5.1	2791	US-08-592-541-133	Sequence 133, Appl
21	33	5.1	2791	US-09-124-698-1	Sequence 1, Appl
22	33	5.1	2791	US-09-124-698-133	Sequence 133, Appl
23	33	5.1	2791	US-09-127-480-1	Sequence 1, Appl
24	33	5.1	2791	US-09-127-480-133	Sequence 133, Appl
25	33	5.1	2791	US-08-496-841C-1	Sequence 1, Appl
26	33	5.1	2791	US-09-124-523-1	Sequence 1, Appl
27	33	5.1	2791	US-09-124-523-133	Sequence 133, Appl

C 28	33	5.1	2792	4	US-08-496-841C-133	Sequence 133, Appl
C 29	33	5.1	3086	3	US-08-888-077A-3	Sequence 3, Appl
C 30	33	5.1	3087	2	US-08-967-101-5	Sequence 5, Appl
C 31	33	5.1	3087	2	US-08-592-541-5	Sequence 5, Appl
C 32	33	5.1	3087	3	US-09-124-698-5	Sequence 5, Appl
C 33	33	5.1	3087	4	US-08-496-841C-5	Sequence 5, Appl
C 34	33	5.1	3087	4	US-08-496-841C-5	Sequence 5, Appl
C 35	33	5.1	3087	4	US-08-496-841C-5	Sequence 5, Appl
C 36	32.5	5.1	87350	3	US-08-781-891-79	Sequence 79, Appl
C 37	32.5	5.1	87543	4	US-09-791-211-3	Sequence 3, Appl
C 38	32.4	5.0	594	4	US-09-385-982-75	Sequence 75, Appl
C 39	32.4	5.0	3282	4	US-09-315-793-51	Sequence 51, Appl
C 40	32.4	5.0	4206	4	US-09-302-620B-81	Sequence 81, Appl
C 41	32	5.0	448	2	US-08-967-101-106	Sequence 106, Appl
C 42	32	5.0	448	2	US-08-592-541-106	Sequence 106, Appl
C 43	32	5.0	448	3	US-09-124-698-106	Sequence 106, Appl
C 44	32	5.0	448	4	US-09-127-480-106	Sequence 106, Appl
C 45	32	5.0	448	4	US-08-496-841C-106	Sequence 106, Appl

ALIGNMENTS

RESULT 1
US-09-222-938A-15
; Sequence 15. Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 344
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-15

Query Match	51.1%	Score	327.8;	DB 4;	Length	344;			
Best Local Similarity	97.9%	Pred. No.	3.1e-95;						
Matches	332;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps	0;
Qy	113	TGTGACCTTGCCTGGTGTGGAGCGGTA	172						
Db	6	TGTGACCTTGCCTGGTGTGGAGCGGTA	65						
Qy	173	ACTAAGTAAAGTTTCCCTGGGATCGTT	232						
Db	66	ACTAAGTAAAGTTTCCCTGGGATCGTT	125						
Qy	233	CGCAAGAGGAGAGTAGAAACCTGAAAC	292						
Db	126	CGCAAGAGGAGAGTAGAAACCTGAAAC	185						
Qy	293	GTCTTAATTTCCACCATTCGACAGCGT	352						
Db	186	GTCTTAATTTCCACCATTCGACAGCGT	245						
Qy	353	GATAGAAGACAGAGATGGAACGCTAATA	412						
Db	246	GATAGAAGACAGAGATGGAACGCTAATA	305						
Qy	413	AAAGTGACAGATCACAGTGTAGCTAAC	451						
Db	306	AAAGTGACAGATCACAGTGTAGCTAAC	344						

RESULT 2

US-09-222-938A-72
; Sequence 72, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222.938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-72

Query Match 15.6%; Score 100; DB 4; Length 582;
Best Local Similarity 52.9%; Pred. No. 1.9e-22;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 TACTAATGCTTATCAAAATAATTAGATCGGATACACACCACTAAGGTAGACAGAG 60
Db 1 TACTAATTAACGAATACTAATAAACCGGATGAACCTATCGTCTGTAAGGTAGACCAAT 60
Qy 61 ACCTAACCTGTTCTAAGAAAATTTAGTTAGATCGCTGCTTACCAAGACCACTTGACCT 120
Db 61 TAACACCGTTCGACCGTTTCTTAACTATAACCCCTCGTGCCTTCGCCCTTGAATCCG 120
Qy 121 TGCTGGTGTGGAAGGCTAAATTCATTTTCGACATACACCTGTTGGAACACTACTG 180
Db 121 CGATGGTTACGTAAGGATGTAAACCACTTTTCGACCAAGCCAGCAGTATCGGCCTTA 180
Qy 181 AAAAGTTTCCTCGGATCGTTGCGACGAAGGCTAATAAAGATAGATGTTCCCAAGA 240
Db 181 TAAACATTTCTCGACCGTTGACGTAACGGAAGAGTACGTACA---ACTATAAGTG 237
Qy 241 GGAGATAGAAACCTGAAACCGCAATAGCGGTATGGAAGGATAGAAACGTCCTTAA 300
Db 238 GCGAAGAAGCTGCTCAGAAACGCAAAATCCGCTGCACAAAGGTAGAACGCTTTAA 297
Qy 301 TTTCCACATTCGACAGCGTTGTCAGGACCTCACTAAAGCCCTAAACGCGGTAGAG 360
Db 298 TTTCCGCAATTCGCGACCGCTGTAGTCTCCGCAAAACGATAAAATGCGTGGGCAAT 357
Qy 361 ACAGATGGGAACGCTAATAGAAGAAACCTCGAGAGATAGAACCGTCTATCACTAAAGTGAC 420
Db 358 AAATAGTCTACCAACGCCATAGAGTAGAGAAATATGAATGATTAAACGAAGAGAG 417
Qy 421 AGATCAGATGCTGCTAGTACGCGGACAATA 452
Db 418 AGTAGCTACAATTTGCCCTAGATATGACAATA 449

RESULT 3

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA

COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 10.5%; Score 67.2; DB 1; Length 7218;
Best Local Similarity 5.3%; Pred. No. 1.8e-11;
Matches 21; Conservative 226; Mismatches 149; Indels 0; Gaps 0;

Qy 158 CATACCGTTGGAACACTACTGAAAGTTTCCCTGGGATCTTCGGACGAGCGTAAT 217
Db 1476 CCTATCTATGCAAGTAGTTAAAGAGATAGAGAAATTTGGTACRRRRRRRRRRRR 1417
Qy 218 AAAAAGTAGATGTTCCGCAAGAGGAGAGTAGAAACCTGAAACCGACAAATAGCCGTAT 277
Db 1416 RR 1357
Qy 278 GGAAGGATAGAACGTCCTAAATTTCCACCATTCGACAGCGTTGGTACAGCCTCACT 337
Db 1356 RRR 1297
Qy 338 AAAAGCCTAAACCGGATAGAACACAGATGGAACGCTAATAGAGAAACCTCGAGAGA 397
Db 1296 RRR 1237
Qy 398 TAGAACCGTCTACTAAAGTGACAGATCACAGTGTGCTAGCTAACGCGCAATAGCCCC 457
Db 1236 RRR 1177
Qy 458 AAGACGAGAAAGTGGAACCAAAATAGGACTCATTTGATCTGAGAGAGAGATAGCGAT 517
Db 1176 RRR 1117
Qy 518 AATAGAACTGTGAACGATCAAACTAATAGTAACGAG 553
Db 1116 RRR 1081

RESULT 4

US-08-463-418-1
; Sequence 1, Application US/08463418
; Patent No. 5908971
; GENERAL INFORMATION:
; APPLICANT: Van Der Straeten, Dominique et al.

```

RESULT 6
US-08-875-972-3/c
; Sequence 3, Application US/08875972
; Patent NO. 5985564
; GENERAL INFORMATION:
; APPLICANT: Huntington Potter and Jinhue Li
; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING
; TITLE OF INVENTION: CHROMOSOME NON-DISJUNCTION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:

```


MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,972
FILING DATE: 08-AUG-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,448
FILING DATE: 16-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Granehan Esq., Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: HU95-03PA
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1222
US-08-875-972-3

Query Match 5.1%; Score 33; DB 2; Length 1488;
Best Local Similarity 55.6%; Pred. No. 0.71;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;
QY 111 ATTGTGACCTGCTGGTGTGGAAGCGGTAATCCATCTTCGACCATACCGTTGGAA 170
DB 1329 ATCTGACTTTGTAGATGGACACAGGAATACCTTTGTCTCCCGAGATTGGTT 1270
QY 171 ACCTAACTGAAAGTTCCCTTGGGATCGTTGCGAGAA---GGCTAATAAAAGTACA 227
DB 1269 ATAGTCAAGAGAACATCCATGGGATCTAACCCCAATATGCTAGATATAAAATTGA 1210
QY 228 TGTCCGCAAGAGGAGAGTAGAAACCTGAAAA 260
DB 1209 TCGAATGCTAATTTGGTCCATAAAAGCGTGTACA 1177

RESULT 7
US-08-743-637B-2
Sequence 2, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2275 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecalis
US-08-743-637B-2
Query Match 5.1%; Score 33; DB 2; Length 2275;
Best Local Similarity 49.7%; Pred. No. 0.86;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 14 AAAATAATTAGGATCGATAGACGACCCCAAGCTAAGGTAGACAGACCTAACCTGTTC 73
DB 1612 AATGGATTTTATGATGACAGACATGCGAGCTATATTGAAAAAATTATCAACCGTTCC 1671
QY 74 ATAAGAAAGTTAGTTAGATCGCTCGTACCAAGACCATTTGTACCTTGTCTGGTTGGA 133
DB 1672 AAAATAAATATATTTTCAGATGCTATTACACGTGTAGACGACGACCAATCAGAAATTAG 1731
QY 134 AGCGGTAAATCATCTTTCCACCATACCGTTGGAAACACTAACTGAA 182
DB 1732 GTGCGCAAGACGGTTTATTCGCCAANTCCGTGAATTACAGGACGCA 1780

RESULT 8
US-08-526-840B-2
Sequence 2, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GEN
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732

FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2275 base pairs
TYPE: nucleic acid
STRANDEDNESS: Double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecalis

US-08-526-840B-2

Query Match 5.1%; Score 33; DB 3; Length 2275;
Best Local Similarity 49.7%; Pred. No. 0.86;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Qy 14 AAATAATTAGTCGGATAGACGACCAAGCTAAGTAGACAGACCTAACTGTTTC 73
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1612 AATGAAATTTGATCAACAAGAATGCGAGCTATATTGAAAAAATTATCAACGGTTTCC 1671
Qy 74 ATAGAAAGTTTACTAGATCGCTCTGACCAAGACCACTTGTGACCTTGCTGGTGTGGA 133
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1672 AAATAAATATTTTACAGTCTATTACACGCTGTAGCAGGACCAATCAGAAATATAG 1731
Qy 134 AGGCTAAATCCATCTTTTGACCATACCTGTTGGAACACATAACTGAA 182
|| || || || || || || || || || || || || || || || || || || || ||
Db 1732 GTGCCCAAGACGGTTTATTGACCAATCCGTGAATTACAGAACGCA 1780

RESULT 9

US-08-923-454A-9/c
Sequence 9, Application US/08923454A
Patent No. 6004794
GENERAL INFORMATION:
APPLICANT: Creasy, Caretha
APPLICANT: Livi, George
APPLICANT: Kartan, Eric
APPLICANT: Clinkenbeard, Helen
APPLICANT: Browne, Michael
APPLICANT: Southan, Christopher
TITLE OF INVENTION: HUMAN SERINE PROTEASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,454A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/025436
FILING DATE: 06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50547
TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2764 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-923-454A-9

Query Match 5.1%; Score 33; DB 3; Length 2764;

Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;
Qy 111 ATTTGACCTTGCCTGGTGTGGAAGCGCTAAATCCATTCTTTGACCATACCTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTAGATGTGCACAGGAAATACCTTTGTCTCCCGAGATTGGTT 1697
Qy 171 ACATTAACCTGAAAAGTTTCTTGGGATCGTTGCGACGAA---GGCTAATAAAAGTAGA 227
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATACTCAAGAAGAACAATCCATGGGATTTCTAACCGCAATATGCTAGATATAAATTTGA 1637
Qy 228 TGTTCGCCAAGAGGAGTAGAAGACCTGAAA 260
|| || || || || || || || || || || || || || || || || || || || ||
Db 1636 TGGAAATGCTAATTTGGTCCATATAAAAGGCTGTACA 1604

RESULT 10

US-08-832-867-2/c
Sequence 2, Application US/08832867C
Patent No. 6376239

GENERAL INFORMATION:

APPLICANT: BAUMEISTER, Ralf
TITLE OF INVENTION: DNA MOLECULES COMPRISING A PROMOTER CAPABLE OF
TITLE OF INVENTION: CONFERRING EXPRESSION OF A HETEROLOGOUS DNA SEQUENCE IN
TITLE OF INVENTION: ALL NEURAL CELLS AT ALL STAGES OF DEVELOPMENT IN C.

FILE REFERENCE: 674503-2004
CURRENT APPLICATION NUMBER: US/08/832,867C
CURRENT FILING DATE: 1997-04-04
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 2

LENGTH: 2764

TYPE: DNA

ORGANISM: Caenorhabditis elegans

US-08-832-867-2

Query Match 5.1%; Score 33; DB 4; Length 2764;

Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

Qy 111 ATTTGACCTTGCCTGGTGTGGAAGCGCTAAATCCATTCTTTGACCATACCTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTAGATGTGCACAGGAAATACCTTTGTCTCCCGAGATTGGTT 1697
Qy 171 ACATTAACCTGAAAAGTTTCTTGGGATCGTTGCGACGAA---GGCTAATAAAAGTAGA 227
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATAGTCAAGAAGAACAATCCATGGGATTTCTAACCGCAATATGCTAGATATAAATTTGA 1637
Qy 228 TGTTCGCCAAGAGGAGTAGAAGACCTGAAA 260
|| || || || || || || || || || || || || || || || || || || || ||
Db 1636 TGGAAATGCTAATTTGGTCCATATAAAAGGCTGTACA 1604

RESULT 11

US-08-888-077A-1/c

; Sequence 1, Application US/08888077A
; Patent No. 6020143
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMWENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE AND USES THEREFOR.
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
; STREET: 600 SOUTH AVENUE WEST
; CITY: WESTFIELD
; STATE: NJ
; COUNTRY: USA
; ZIP: 07090-1497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,077A
; FILING DATE: 03-JUL-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,541
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: PALISI, THOMAS M
; REGISTRATION NUMBER: 36,629
; REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP IV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 654-5000
; TELEFAX: (908) 654-7866
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2675
; OTHER INFORMATION: /note= "hps1-1"
; US-08-888-077A-1

Query Match 5.1%; Score 33; DB 3; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;
Qy 111 ATTGTGACCTTGCTGGTGTGGAGGCGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTAGATGGACACAGGAATACACCTTGTCTCCCGAGATTGGTT 1697
Qy 171 ACACCTACTGAAAGTTTCTTGGGATCGTTGGAGCAA---GGCTAATAAAAGTAGA 227
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATAGTCAAGAAGAAACATCGGATTCCTAACCCGAAATATGCTAGATATAAAATTGA 1637
Qy 228 TGTTCGCAAGAGGAGGAGTAGAAACCTGAAAA 260
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

RESULT 12
US-08-706-344C-1/c
; Sequence 1, Application US/08706344C
; Patent No. 6248555
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH

; APPLICANT: WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,344C
; FILING DATE: 30-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/003,054
; FILING DATE: 31-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U.
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0609.4180001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; US-08-706-344C-1

Query Match 5.1%; Score 33; DB 4; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;
Qy 111 ATTGTGACCTTGCTGGTGTGGAGGCGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTAGATGGACACAGGAATACACCTTGTCTCCCGAGATTGGTT 1697
Qy 171 ACACCTACTGAAAGTTTCTTGGGATCGTTGGAGCAA---GGCTAATAAAAGTAGA 227
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATAGTCAAGAAGAAACATCGGATTCCTAACCCGAAATATGCTAGATATAAAATTGA 1637
Qy 228 TGTTCGCAAGAGGAGGAGTAGAAACCTGAAAA 260
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

RESULT 13
US-08-706-344C-3/c
; Sequence 3, Application US/08706344C
; Patent No. 6248555
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; APPLICANT: WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON

```
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,344C
FILING DATE: 30-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/003,054
FILING DATE: 31-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: KIM, JUDITH U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 0609.4180001
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2765 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 249..1649
US-08-706-344C-3

Query Match 5.1%; Score 33; DB 4; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

OY 111 ATTGTGACCTGCTGTTGGAGCGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
DB 1756 ATCTTGACCTGTTAGATGTGCACACAGGAAATCACCTTTGCTCTCCCGAGATTGGTT 1697

OY 171 ACACCTAACTGAAAAGTTTCTTGGGATCGTTGGACGAA---GGCTAATAAAAGTAGA 227
DB 1696 ATAGTCAAGAGAAACATCCATGGATTCTAACCGCAAAATATGCTAGATATAAAATTGA 1637

OY 228 TGTTCGCAAGAGAGAGTAGAACCTGAAAA 260
DB 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

RESULT 14
US-08-706-344C-27/c
; Sequence 27, Application US/08706344C
; Patent No. 6248555
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; APPLICANT: WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,344C
FILING DATE: 30-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/003,054
FILING DATE: 31-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: KIM, JUDITH U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 0609.4180001
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2765 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 249..1649
US-08-706-344C-3

Query Match 5.1%; Score 33; DB 4; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

OY 111 ATTGTGACCTGCTGTTGGAGCGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
DB 1756 ATCTTGACCTGTTAGATGTGCACACAGGAAATCACCTTTGCTCTCCCGAGATTGGTT 1697

OY 171 ACACCTAACTGAAAAGTTTCTTGGGATCGTTGGACGAA---GGCTAATAAAAGTAGA 227
DB 1696 ATAGTCAAGAGAAACATCCATGGATTCTAACCGCAAAATATGCTAGATATAAAATTGA 1637

OY 228 TGTTCGCAAGAGAGAGTAGAACCTGAAAA 260
DB 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

RESULT 15
US-08-706-344C-29/c
; Sequence 29, Application US/08706344C
; Patent No. 6248555
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; APPLICANT: WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
```

```

; NAME: KIM, JUDITH U.
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0609.4180001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; US-08-706-344C-29

```

```

Query Match      5.1%; Score 33; DB 4; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

Qy 111 ATTGTGACCTTGCTGGTTGTGGAAGGCGTAAATCCATTTTCGACCATACCGTTGGAA 170
   || || || || || || || || || || || || || || || || || || || ||
Db 1756 ATCTTGACTTTGTTAGATGTGCACACAGGAAATCACCTTTGTCTCCCCAGATTGGTT 1697

Qy 171 ACACCTAACTGAAAAGTTTCCTTGGATCGTTGCGACGAA---GGCTAATAAAAGTAGA 227
   || || || || || || || || || || || || || || || || || || || ||
Db 1696 ATAGTCAAGAGAAGAACATCCATGGGATTCCTAACCGCAAAATATGCTAGATATAAAATTGA 1637

Qy 228 TGTTCGCAAGAGGAGAGTAGAAGAACCTGAAAA 260
   || || || || || || || || || || || || || || || || || || || ||
Db 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

```

Search completed: December 25, 2002, 14:26:36
Job time : 110.625 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 : Search time 189.61 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgttatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues

Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published_Applications_NA.*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/2/pubpna/PT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/2/pubpna/PTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	642	100.0	642	12	US-10-068-080-11
2	100	15.6	582	12	US-10-068-080-12
C 3	35.6	5.5	640681	10	US-09-790-988-1
C 4	34.2	5.3	180216	10	US-09-835-232-6
C 5	33.6	5.2	1058	9	US-09-938-842A-2917
6	33.4	5.2	236	10	US-09-864-761-33009
7	33.4	5.2	550	10	US-09-864-761-16513
8	33.4	5.2	29729	10	US-09-070-927A-238
9	33	5.1	1155	10	US-09-815-242-6498
10	33	5.1	2275	10	US-09-452-599-2
C 11	33	5.1	2765	10	US-09-785-474-1
C 12	33	5.1	2765	10	US-09-785-474-3
C 13	33	5.1	2765	10	US-09-785-474-27
C 14	33	5.1	2765	10	US-09-785-474-29
C 15	33	5.1	2765	10	US-09-785-474-31
C 16	32.6	5.1	691	10	US-09-878-574-4162
C 17	32.6	5.1	1125	10	US-09-974-300-4278
C 18	32.6	5.1	2744	10	US-09-070-927A-548
C 19	32.6	5.1	1503841	9	US-09-946-807-1

C 20	32.6	5.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 21	32.6	5.1	1503841	10	US-09-795-686-1	Sequence 1, Appli
C 22	32.4	5.0	921	10	US-09-974-300-4778	Sequence 4778, Ap
C 23	32.4	5.0	1560	9	US-09-764-868-1432	Sequence 1432, Ap
C 24	32.4	5.0	4206	10	US-09-911-781-2	Sequence 2, Appli
C 25	32.4	5.0	7258	10	US-09-790-988-3	Sequence 3, Appli
C 26	32.2	5.0	396	10	US-09-878-574-1207	Sequence 1207, Ap
C 27	32.2	5.0	615	9	US-09-950-933A-28	Sequence 28, Appli
C 28	32.2	5.0	684973	10	US-09-263-959-1	Sequence 1, Appli
C 29	32	5.0	541	10	US-09-864-781-14853	Sequence 14853, A
C 30	32	5.0	2000	9	US-09-938-842A-4002	Sequence 4002, Ap
C 31	31.8	5.0	487	9	US-10-046-935-701	Sequence 701, App
C 32	31.8	5.0	487	9	US-09-878-178-701	Sequence 701, App
C 33	31.8	5.0	2000	9	US-09-938-842A-4834	Sequence 4834, Ap
C 34	31.6	4.9	31885	10	US-09-764-860-775	Sequence 775, App
C 35	31.6	4.9	31885	10	US-09-764-877-2530	Sequence 2530, Ap
C 36	31.6	4.9	31885	10	US-09-764-877-2541	Sequence 2541, Ap
C 37	31.4	4.9	493	10	US-09-895-035-3	Sequence 3, Appli
C 38	31.2	4.9	202	10	US-09-764-869-159	Sequence 159, App
C 39	31.2	4.9	464	10	US-09-783-590-9045	Sequence 9045, Ap
C 40	31.2	4.9	1968	9	US-09-991-496-130	Sequence 130, App
C 41	31.2	4.9	2000	9	US-09-938-842A-4256	Sequence 4256, Ap
C 42	31	4.8	385	10	US-09-920-300A-472	Sequence 472, App
C 43	31	4.8	385	12	US-10-033-528-472	Sequence 472, App
C 44	31	4.8	392	10	US-09-867-701-1908	Sequence 1908, Ap
C 45	31	4.8	440	10	US-09-998-598-1083	Sequence 1083, Ap

ALIGNMENTS

RESULT 1
US-10-068-080-11
; Sequence 11, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-11

Query Match	100.0%	Score 642;	DB 12;	Length 642;
Best Local Similarity	100.0%;	Pred. No. 3.8e-181;		
Matches 642;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	TACTAATGTTATCAAAATATTAGGATCGATAGACAGACCCAGCTAAGGTAGACAGAG 60			
Db 1	TACTAATGTTATCAAAATATTAGGATCGATAGACAGACCCAGCTAAGGTAGACAGAG 60			
QY 61	ACCTAACCTGTTTCATAGAAGTTTATGTTAGATCGCTCGTACCAAGACCATTTGTGACCT 120			
Db 61	ACCTAACCTGTTTCATAGAAGTTTATGTTAGATCGCTCGTACCAAGACCATTTGTGACCT 120			
QY 121	TGCTGTTGTTGGAAGCGGTAATAATCCATTTCTTCGACCATACCGTTGGAAACACTAACGTG 180			
Db 121	TGCTGTTGTTGGAAGCGGTAATAATCCATTTCTTCGACCATACCGTTGGAAACACTAACGTG 180			
QY 181	AAAAGTTTCTTGGGATCGTTCCGACGAGGCTAATAAAGTAGATGTTCCGGCAAGA 240			
Db 181	AAAAGTTTCTTGGGATCGTTCCGACGAGGCTAATAAAGTAGATGTTCCGGCAAGA 240			
QY 241	GGAGACTAGAAACCTCTAAACCCAGCAATAGCCGGTGTGGAAGGATAGAAACGCTCTATA 300			


```

; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33009
; LENGTH: 236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008166.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: NT HIT: AL163245.2, EVALU6 6.00e-41
; OTHER INFORMATION: EST_HUMAN HIT: AI436093.1, EVALU6 4.00e-43
; OTHER INFORMATION: SWISSPROT HIT: P43679, EVALU6 3.90e-01
; -09-864-761-33009

Query Match          5.2%; Score 33.4; DB 10; Length 236;
Best Local Similarity 51.7%; Pred.No. 1.2;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0

QY 430 TGTCTAGCTAACGCCGACAAATAGCCCCAGACGAGAAAAGGTGAACCAATAGGAC 489
      ||||| |||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 TGTCAGACTCTAACGGTGCGGAACCACAGTCACCTCAACTACAAAAATTAAATACAGTGGAA 117

QY 490 TCATTGTTACTGAGACAGAAGTAGCCCATATAGCAANTCGTGACGATCAAACTAATAGTAA 549
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 TAATTGGATCCCGAGCAGCAGAACGCCAAGTGGTGGCACTGAACCAATAAATCAAGGTGA 177

QY 550 GCAGTATTCTCTGTTATATCGAGCATAG 576
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 GCATAGCTACTGTAATAAACAGCAGAG 204

RESULT 7
US-09-864-761-16513
; Sequence 16513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
```



```

US-09-070-927A-238
; Sequence 238, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptide
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 238:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; .SEQUENCE DESCRIPTION: SEQ ID NO: 238:
US-09-070-927A-238

Query Match          5.2%; Score 33.4; DB 10; Length 29729;
Best Local similarity 49.7%; Pred. No. 10;
Matches      85; Conservative    0; Mismatches 86; Indels    0; Gaps    0;

Qy   12 TCAATAATTTAGGATCGGTACAGACCACCAAGCTAAGGTAGACACGACCTACACTTGT 71
     || || |||| ||| ||| || | || | || | || | || | || | || | || |
Db   3461 TAAATGAATTTTGATGGAACAAGAATCATGCACCTATATTGAAAAAATTTATTCAAACGTTT 3520
     || || || || || || || || || || || || || || || || || || || ||
Qy   72 TCATAGAAAAGTTTAGTTAGATCGCTCGCTACCAAGACCACTTGTGACCTTGGCTGTTGFG 131
     || || || || || || || || || || || || || || || || || || || ||
Db   3521 CCATAATAATATATTTTCAGATGCTATTACAGTGTAGCAGCCGACCAATCAGAAATTT 3580
     || || || || || || || || || || || || || || || || || || || ||
Qy   132 GAAGCGGTAAAAATCCATTCTTTGACACCATCCGTTGGAAAACTAACTGAA 182
     || || || || || || || || || || || || || || || || || || || ||
Db   3581 AGGTGCGCAAGACGGTTTATTTCGACCAATCCGTGAATTACAGGAACGCCAA 3631
     || || || || || || || || || || || || || || || || || || || ||

RESULT 9
US-09-815-242-6498
; Sequence 6498, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
```

```

3461 TAAATGGAATTTCAGTGGAAACCAAGCAAGCCCTATTATTGAAAAAATAATTTCACCGTTT 3520
QY      72 TCATAPAGAAGTTTAGTTAGATGCCTCGTGACCAGACCATTTGTGACCTTGCTGCTGTGTG 131
Dk      ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
3521 CCAPAATAAATAATATATTTTCAGATGCTATTACACGTGTGTAGCACGGACACCAATTCAGAAAATT 3580
QY      132 GAACGCGTAAAAATCCATCTCTTCACCATACATACCTGTGGAAACACTTAAGTGAA 182
Dk      |||| | ||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
3581 AGGTGCGCAAGACGGTTTTATTTCACCAATCCGTGAATTACAGGAACGCCAA 3631
RESULT 9
US-09-815-242-6498
; Sequence 6498, Application US/09815242
; Patent No. US20020051369A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.

```

; APPLICANT: OHlsen, Karl L.
; APPLICANT: Zyskind, Judith W.

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6498
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1155)
US-09-815-242-6498

Query Match      5.1%; Score 33; DB 10; Length 1155;
Best Local Similarity 49.7%; Pred. No. 3.2;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Oy  14 AAATAATTAGATCGGATAGACAGCAACCAAGCTAAGGTAGACACGACCTAACTGTTTC 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  779 AATGGAATTTTGTATGAACAAGCAATCGACCTATATTGAAAAAATTATTCAACGTTTCC 838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy  74 ATAAGAAAGTTTAGTTAGATCGCTGCTACCAAGACCAATTTGACCTTGTGTTGGGA 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  839 AAATAAATATATTTCAGATCTATTACACGTGTAGCAGCAGCAATCAGAAAAATTAG 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy  134 AGCGTAAATCCATTCTTTTCGACCATACACCTGTTGGAACACTAACTGAA 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  899 GTGCGCAAGACGGTTTATTTCGACCAATCCGTTGAATTACAGGAACGCAA 947
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-452-599-2
; Sequence 2, Application US/09452599
; Patent No. US20020055101A1
; GENERAL INFORMATION:
; APPLICANT: Bergeron, Michel G.
; APPLICANT: Ouellette, Marc
; APPLICANT: Roy, Paul H.
; TITLE OF INVENTION: Specific and Universal Probes and Amplification Primers
; TITLE OF INVENTION: to Rapidly Detect and Identify Common Bacterial
; TITLE OF INVENTION: Pathogens and Antibiotic Resistance Genes from Clinical
; TITLE OF INVENTION: Specimens for Routine Diagnosis in Micro
; FILE REFERENCE: 12287.31
; CURRENT APPLICATION NUMBER: US/09/452,599
; CURRENT FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: 08/526,840
; PRIOR FILING DATE: 1995-09-11
; PRIOR APPLICATION NUMBER: 08/304,732
; PRIOR FILING DATE: 1994-09-12
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2
; LENGTH: 2275
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-452-599-2

Query Match      5.1%; Score 33; DB 10; Length 2275;
Best Local Similarity 49.7%; Pred. No. 4.3;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Oy  14 AAATAATTAGATCGGATAGACAGCAACCAAGCTAAGGTAGACACGACCTAACTGTTTC 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1612 AATGGAATTTTGTATGAACAAGCAATCGACCTATATTGAAAAAATTATTCAACGTTTCC 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy  74 ATAAGAAAGTTTAGTTAGATCGCTGCTACCAAGACCAATTTGACCTTGTGTTGGGA 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1672 AAATAAATATATTTCAGATCTATTACACGTGTAGCAGCAGCAATCAGAAAAATTAG 1731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy  134 AGCGTAAATCCATTCTTTTCGACCATACACCTGTTGGAACACTAACTGAA 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1732 GTGCGCAAGACGGTTTATTTCGACCAATCCGTTGAATTACAGGAACGCAA 1780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-785-474-1/c
; Sequence 1, Application US/09785474
; Patent No. US20010012626A1
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; APPLICANT: WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/09/785,474
; APPLICATION NUMBER: 08/706,344
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: 60/003,054
; FILING DATE: 31-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U.
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0609.4180002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-785-474-1
```

Query Match 5.1%; Score 33; DB 10; Length 2765;
 Best Local Similarity 55.6%; Pred. No. 4.7;
 Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTTGGTGTGGAAGCGGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1756 ATCTTGACTTTGTAGATGTGGACACAGGAAATCACCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAAGTAAAGTTTCTTGGGATCGTTGCGACGAA---GCCTAATAAAGTAGA 227
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1696 ATAGTCAAGAAGAAACATCCATGGGATCTTAACCGCAATATGCTAGATATAAAATTGA 1637

QY 228 TGTTCGCCAAGAGGAGAGTAGAAACCTGAAAA 260
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1636 TGAATGCTAATTGGTCCATATAAAGGCTGTACA 1604

RESULT 12

US-09-785-474-3/c
 ; Sequence 3, Application US/09785474
 ; Patent No. US20010012626A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TANZI, RUDOLPH
 ; WASCO, WILMA
 ; TITLE OF INVENTION: Genetic Alterations Related To Familial
 ; Alzheimer's Disease
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 NEW YORK AVENUE, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/785,474
 ; FILING DATE: 20-Feb-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/706,344
 ; FILING DATE: 30-AUG-1996
 ; APPLICATION NUMBER: 60/003,054
 ; FILING DATE: 31-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KIM, JUDITH U.
 ; REGISTRATION NUMBER: 40,679
 ; REFERENCE/DOCKET NUMBER: 0609.4180002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-2600
 ; TELEFAX: 202-371-2540
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2765 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 249..1649
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 US-09-785-474-3

Query Match 5.1%; Score 33; DB 10; Length 2765;
 Best Local Similarity 55.6%; Pred. No. 4.7;
 Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTTGGTGTGGAAGCGGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1756 ATCTTGACTTTGTAGATGTGGACACAGGAAATCACCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAAGTAAAGTTTCTTGGGATCGTTGCGACGAA---GGCTAATAAAGTAGA 227
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1696 ATAGTCAAGAAGAAACATCCATGGGATCTTAACCGCAATATGCTAGATATAAAATTGA 1637

QY 228 TGTTCGCCAAGAGGAGAGTAGAAACCTGAAAA 260
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1636 TGAATGCTAATTGGTCCATATAAAGGCTGTACA 1604

RESULT 13
 US-09-785-474-27/c
 ; Sequence 27, Application US/09785474
 ; Patent No. US20010012626A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TANZI, RUDOLPH
 ; WASCO, WILMA
 ; TITLE OF INVENTION: Genetic Alterations Related To Familial
 ; Alzheimer's Disease
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 NEW YORK AVENUE, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/785,474
 ; FILING DATE: 20-Feb-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/706,344
 ; FILING DATE: 30-AUG-1996
 ; APPLICATION NUMBER: 60/003,054
 ; FILING DATE: 31-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KIM, JUDITH U.
 ; REGISTRATION NUMBER: 40,679
 ; REFERENCE/DOCKET NUMBER: 0609.4180002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-2600
 ; TELEFAX: 202-371-2540
 ; INFORMATION FOR SEQ ID NO: 27:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2765 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 249..1649
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 27:
 US-09-785-474-27

Query Match 5.1%; Score 33; DB 10; Length 2765;
 Best Local Similarity 55.6%; Pred. No. 4.7;
 Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTTGGTGTGGAAGCGGTAAATCCATTCTTTCGACCATACCGTTGGAA 170
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1756 ATCTTGACTTTGTAGATGTGGACACAGGAAATCACCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAAGTAAAGTTTCTTGGGATCGTTGCGACGAA---GGCTAATAAAGTAGA 227

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 24, 2002, 22:13:55 ; Search time 10196.6 seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-12
Perfect score: 582
Sequence: 1 tacaattaaacgaataacta.....gatttcattttaccaatatt 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	49	8.4	4574	9	HSM803437	AL832130 Homo sapi
2	48	8.2	7218	6	I66494	I66494 Sequence 14
3	47.6	8.2	152103	2	AC012130	AC012130 Homo sapi
4	47.6	8.2	186609	2	AC083872	AC083872 Homo sapi
5	46.6	8.0	12025	6	AX346200	AX346200 Sequence
6	46	7.9	24175	2	AC115594	AC115594 Dictyoste
7	45.4	7.8	183345	2	AC120550	AC120550 Mus muscu
8	45.4	7.8	203422	2	AC107857	AC107857 Mus muscu
9	45.2	7.8	17480	3	AC114258	AC114258 Dictyoste
10	44.8	7.7	105470	2	AC116306	AC116306 Dictyoste
11	44.4	7.6	6641	6	AX281294	AX281294 Sequence
12	44.4	7.6	6641	6	AX345217	AX345217 Sequence
13	44.4	7.6	120571	9	CNS01073	AL132633 Human chr
14	44.4	7.6	172311	2	AC021729	AC021729 Homo sapi
15	44.4	7.6	218956	9	CNS00008	AL049831 Human chr
16	44.2	7.6	124820	2	AC117073	AC117073 Dictyoste
17	44.2	7.6	168543	2	AL845301	AL845301 Danio rer
18	44	7.6	58894	9	AC093277	AC093277 Homo sapi
19	44	7.6	170392	9	AC060765	AC060765 Homo sapi
20	44	7.6	180697	9	AC114977	AC114977 Homo sapi
21	43.8	7.5	152409	2	PFMALIP1	AL031744 Plasmodiu
22	43.6	7.5	100920	9	AC020921	AC020921 Homo sapi
23	43.6	7.5	167872	2	AC007554	AC007554 Homo sapi
24	43.6	7.5	251664	2	AC008908	AC008908 Homo sapi
25	43.4	7.5	86827	3	PFMALIP5	AL034556 Plasmodiu
26	43	7.4	158376	9	AC099059	AC099059 Homo sapi
27	43	7.4	194281	2	AC078812	AC078812 Homo sapi
28	42.8	7.4	69048	2	AC011826	AC011826 Homo sapi
29	42.6	7.3	131201	2	AC130820	AC130820 Mus muscu
30	42.6	7.3	177604	2	AC013342	AC013342 Homo sapi
31	42.6	7.3	183338	2	AC083775	AC083775 Homo sapi
32	42.6	7.3	188014	9	AC093744	AC093744 Homo sapi
33	42.2	7.3	1852	9	BC028121	BC028121 Homo sapi
34	42.2	7.3	57835	2	AC110290	AC110290 Homo sapi
35	42	7.2	77835	2	PFMALIP2_3	Continuation (4 of
36	42	7.2	168469	9	AC090592	AC090592 Homo sapi
37	42	7.2	168482	2	AC025235	AC025235 Homo sapi
38	42	7.2	177628	2	AL772143	AL772143 Danio rer
39	41.3	7.2	2863	3	EHAVAMER	X75436 E.histolyti
40	41.3	7.2	2913	3	ENHRRSA	M95498 Entamoeba h
41	41.3	7.2	3188	3	ENHRRSA	L36807 Entamoeba h
42	41.3	7.2	30726	2	AC117269	AC117269 Dictyoste
43	41.3	7.2	65691	3	PFMALIP1	Z97348 Plasmodium
44	41.3	7.2	98734	2	PFMALIP2	AL031745 Plasmodiu
45	41.3	7.2	106216	9	AC061997	AC061997 Homo sapi

ALIGNMENTS

RESULT 1
HSM803437
LOCUS Homo sapiens mRNA; cDNA DKF2p666H1210 (from clone DKF2p666H1210).
DEFINITION 4574 bp mRNA linear PRI 12-JUL-2002
ACCESSION AL832130
VERSION AL832130.1 GI:21732673
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4574)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de.
Resequenced by ACOWA (Berlin/Germany) within the CDNA sequencing consortium of the German Genome Project.
This clone (DKFZp686H1210) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

```

FEATURES
  source
    1. .4574
      Location/Qualifiers
        organism="Homo sapiens"
        db_xref="taxon:9606"
        /clone="DKFZp686H1210"
        /tissue_type="cDNA-collection"
        /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
          DH10B; sites SfiIA + SfiIB"
        /dev_stage="adult"
      polyA_signal 4394..4399
      polyA_site 4409
      BASE COUNT 1349 a 1085 c 1074 g 1066 t
      ORIGIN

```

Query Match 8.4%; Score 49; DB 9; Length 4574;
Best Local Similarity 47.5%; Pred. No. 0.038;
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy	213	AAAGAGTACGTACAACACTATAAGTGGCGGAAGAACGTCCTCAGAAACGCCCAAAATCCGGT	272
Db	4270	AAAAAAAAAAAAAAAAAATCCCTGAATGATGATTAGAGACATCACCGCTAAAAAACAATAC	4329
Qy	273	GCACAAGGGTGAAGCGGGTTTAAATTTCCGGCCATTTCCGGCACCGCTGTAGTCTCTCCGCA	332
Db	4330	ATTATATAAGCTAGGATTTGTGTATATGCAAAATATTTCCTGCCTCTCTTTTCTCTGTTTA	4389
Qy	333	AAACGATAAATTCGCTGGGCGACAATAAATAGTGTACCAACGCCCATAGAGAGTAGAAAAA	392
Db	4390	AAACATAAATAGCGATTTGTATAAAAAANAANAANAANAANAANAANAANAANAANAANA	4449
Qy	393	TATGAACTCGATTTAAACAAAGAGAGAGTAGCTACAAATTTGCCCTAGATATGACACAATATAT	452
Db	4450	AAAAAAAAAAAAAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	4509
Qy	453	ATCAAAAGAAAGACAGGTACTATGTCATAAATAACTTAACAGCAATGGCAGAGTCGATAAAAAACA	512
Db	4510	AAAAAAAAAAAAAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	4569
Qy	513	CTATA	517
Db	4570	AAAAA	4574

RESULT 2					
LOCUS	I66494/C				
DEFINITION	I66494		7218 bp	DNA	linear
ACCESSION	Sequence 14 from patent	US 5670367.			
VERSION	I66494.1				
KEYWORDS	GI:2724471				
SOURCE	.				
	Unknown.				

REFERENCE 1 (bases 1 to 7218)
 AUTHORS Dörner F., Scheiflinger F. and Falkner, F. Gunter.
 TITLE Recombinant fowlpox virus
 JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
 FEATURES Location/Qualifiers
 source 1. .7218

BASE COUNT	1944 a	19491 c	1486 g	1929 t	368 others
ORIGIN					

Query Match . 8.28; Score 48; DB 6; Length 7218;

[illegible][illegible]

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152103)

REFERENCE
AUTHORS
Bliren, B., Linton, L., Nusbaum, C. and Lander, E.

2 (bases 1 to 152103)

Blirren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgalter, B., C-own, A., Castle, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gholand, J., Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lahoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Tefameg, J., Tefameg, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 2000 this sequence version replaced gi:6087954.
All repeats were identified using RepeatMasker:
Snit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: L2823

Qy 381 GAAGTAGAAAAATATGAAGTCAATTAAACAAGAGAGTAGTACAAATGTGTCCTAGAT 440
| | | | |
Db 5144 GGAGGACAAAATCCATTTGCCATGTGAAGAATAAGACAGAAGCCTGACTGTGCCTCTCA 54203

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part

of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

ECORI BglII HindIII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	8908	9428	9477	1091	1108
6	<800	2067	2034	6382	6457
1282	1295	5146	5237	512	<800
5875	5795	214	<800	449	<800
864	875	6664	6551	2332	2341
3210	3260	2112	2139	3406	3432
547	<800	558	<800	4842	4741
2719	2686	2175	2139	341	<800
12	<800	2550	2568	6269	6086
3603	3651	2441	2485	6163	6086
4317	4277	8416	8390	2196	2194
579	<800	4875	4728	171	<800
4039	4057	1489	1456	939	957
8217	8202	15165	15187	679	<800
1539	1508	3130	3275	4331	4364
371	<800	1521	1456	618	<800
16481	16182	6483	6345	5175	5188
695	<800	1151	1140	1062	1108
1217	1182	609	<800	1506	1461
729	721	4775	4728	4663	4741
7979	8202	3602	3532	2141	2194
1517	1508	282	<800	275	<800
304	<800	1257	1264	9412	9422

5107	6102	10	<800	1133	1108
11972	12030	2735	2848	1290	1261
893	875	3563	3532	5872	6086
828	875	11001	11223	1949	1916
1149	1182	4671	4728	1430	1461
5513	5513	1494	1456	2544	2595
8138	8202	5095	5079	2555	2595
131	<800	3991	3944	6162	6086
1346	1295	1450	1456	2058	2194
3683	3651	2231	2267	6346	6457
285	<800	1043	1031	41	<800
480	<800	11317	11223	3020	2972
10788	10771	4313	4310	56	<800
12229	12030	1819	1789	6425	6457
10595	10339	2613	2678	474	<800
1804	1800	5863	5777	4380	4364
19511	19785	1410	1456	2187	2194
6130	6102	4258	4111	3234	3266
5058	5030	235	<800	1460	1461
262	<800	1484	1456	5208	5188
4871	4887	209	<800	3224	3266
5728	5795	539	<800	161	<800
3009	3046	1298	1264	2313	2341
2982	3061	2904	2972		
6198	6120	3309	3266		
871	862	601	<800		
2286	2387	124	<800		
457	<800	14221	14186		
3725	3698	6344	6086		
432	<800	2218	2194		
2926	2968	4420	4364		
1590	1662	3743	3764		
70	<800	4425	4741		
333	<800	724	<800		
274	<800	247	<800		

Query Match 8.2%; Score 47.6; DB 9; Length 186609;
Best Local Similarity 52.5%; Pred. No. 0.088;
Matches 104; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (08-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 183345)

REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 9, 2002 this sequence version replaced gi:20503162.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20738
Center clone name: 142.I.8
----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173947 bases at least Q40
Consensus quality: 175704 bases at least Q30
Consensus quality: 176645 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 180745; sum-of-contigs
Quality coverage: 7.9 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 183: contig of 183 bp in length
184 283: gap of 100 bp
284 982: contig of 699 bp in length
983 1082: gap of 100 bp
1083 1790: contig of 708 bp in length
1791 1890: gap of 100 bp
1891 2497: contig of 607 bp in length
2498 2597: gap of 100 bp
2598 3293: contig of 696 bp in length
3294 3393: gap of 100 bp
3394 4051: contig of 658 bp in length
4052 4151: gap of 100 bp
4152 4811: contig of 660 bp in length
4812 4911: gap of 100 bp
4912 34646: contig of 29735 bp in length
34647 34746: gap of 100 bp
34747 35900: contig of 1154 bp in length
35901 36000: gap of 100 bp
36001 37699: contig of 1699 bp in length
37700 37799: gap of 100 bp
37800 38952: contig of 1153 bp in length
38953 39052: gap of 100 bp
39053 41334: contig of 2282 bp in length
41335 41434: gap of 100 bp
41435 43398: contig of 1964 bp in length
43399 43498: gap of 100 bp
43499 45252: contig of 1754 bp in length
45253 45352: gap of 100 bp
45353 48902: contig of 3550 bp in length
48903 49002: gap of 100 bp
49003 52020: contig of 3018 bp in length
52021 52120: gap of 100 bp
52121 57368: contig of 5248 bp in length
57369 57468: gap of 100 bp
57469 65330: contig of 7862 bp in length
65331 65430: gap of 100 bp
65431 74541: contig of 9111 bp in length
74542 74641: gap of 100 bp
74642 85758: contig of 11117 bp in length
85759 85858: gap of 100 bp
85859 96006: contig of 10148 bp in length
96007 96106: gap of 100 bp
96107 108889: contig of 12783 bp in length
108890 108989: gap of 100 bp
108990 119894: contig of 10905 bp in length
119895 119994: gap of 100 bp
119995 134363: contig of 14369 bp in length
134364 134463: gap of 100 bp
134464 154899: contig of 20436 bp in length
154900 154999: gap of 100 bp
155000 183236: contig of 28237 bp in length
183237 183336: gap of 100 bp
183337 183345: contig of 9 bp in length.

FEATURES
Location/Qualifiers
1. .183345
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-142I8"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .183
/note="assembly_fragment"
284. .982
/note="assembly_fragment"
1083. .1790
/note="assembly_fragment"
1891. .2497
/note="assembly_fragment"

misc_feature
misc_feature
misc_feature
misc_feature

```

misc_feature      2598..3293
/Note="assembly_fragment"
3394..4051
/Note="assembly_fragment"
4152..4811
/Note="assembly_fragment"
4912..34646
/Note="assembly_fragment"
34747..35900
/Note="assembly_fragment"
36001..37699
/Note="assembly_fragment"
37800..38952
/Note="assembly_fragment"
39053..41334
/Note="assembly_fragment"
41435..43398
/Note="assembly_fragment"
43499..45252
/Note="assembly_fragment"
45353..48902
/Note="assembly_fragment"
49003..52020
/Note="assembly_fragment"
52121..57368
/Note="assembly_fragment"
57469..65330
/Note="assembly_fragment"
65431..74541
/Note="assembly_fragment"
74642..85758
/Note="assembly_fragment"
85859..96006
/Note="assembly_fragment"

Query Match      7.8%; Score 45.4; DB 2; Length 183345;
Best Local Similarity 49.4%; Pred. No. 0.33;
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

OY 332 AAAACGATTAATTCGTTGGGACAAATAATAGTCTACCAACCCCATAGAGTAGAGAA 391
    ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6437 AATAAATAAATTTCAAAAAGAAATATATATATATATATATATATATATATATATAT 6496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 392 ATATGAACATGATTTAAACAAGACAGAGTAGCTACAAATGTCCCTAGATGACATATA 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6497 ATATATATATATATATATACAGCAAGATGGGTGTTGGTAGGGCAATAGAGAAAGTG 6556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 452 TATCAAGAAACAGGTAATGATGATCAATAAATACTAACAGCAATGGGACGAGTGATAAAAC 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6557 TGTATATCTAAGAGTAGTGACTGTAGAAATACTAACACAAAATAAAAGTAAATCTC 6616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 512 ACTATATGCTGTGGCTCGGCTGTTAATTTGCTTAATAGTATTTTCTCTGGAATTCAT 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6617 ACTTAACCATGACTGCTCATTTAAGACTAGAAATTTATCATGTATCTTGTATTTAT 6675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AC107857
LOCUS      AC107857      203422 bp      DNA      linear      HTG 26-APR-2002
DEFINITION Mus musculus clone RP23-18K24, WORKING DRAFT SEQUENCE, 12 ordered
            pieces.
ACCESSION  AC107857
VERSION    AC107857.3 GI:20330954
KEYWORDS  HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Mus musculus
            Mus musculus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 203422)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus, clone RP23-18K24
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 203422)

AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J.J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyne,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,L., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 203422)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collimore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyne,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Major,J., Marquis,N., Matthews,C.,
McClean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 26, 2002 this sequence version replaced gi:20258511.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20637
Center clone name: 18_K24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197973 bases at least Q40
Consensus quality: 200612 bases at least Q30
Consensus quality: 201671 bases at least Q20
Insert size: 199000; agarose-fp
Insert size: 202322; sum-of-contigs
Quality coverage: 6.3 in Q20 bases; agarose-fp

```

Query Match 7.8%; Score 45.4; DB 2; Length 203422;
Best Local Similarity 49.4%; Pred. No. 0.33;

PGSSSIFGICPGFDYQOHHYFTLLSAPLIVVLIISTVFFYIVYVYKLIQKLNKI
 IRBOEDIKSISYSSIIIVSDPGTINSSLOKSLERKSLRKQOQOQOPLLEND
 DSDSFSTNTTNDGDLSDRQSTYIADGGRHSILGKSISTSGSLSSOVTLLDNATAI
 PSNQTLSOOLGAFIKMLNITPICKATFELPCFENGVGSAYVLIANGVSCYTDQY
 HKALVYSCLLIIVIGFPLIMILLFNKRSDDPHQTYIGVILKYSFFWVDVI
 LLRLVLLITMSDPTSAARFLLVSVLVYLLQKYOPEKRIISDRLLELTSLLGK
 FICVYLDNTVQSIQWIVISSIIIFLHAYVILSYNIOIISDLOHLYLXISRGK
 KGRKSNNTLFGSNYFLYNQHHYNDODFSRDSGIFSKKDKKQSRDITVKVSKINN
 NNENNTENNENNENNENNENNVKETEILEESKTFPELNENNYDDSDDFDEDES
 DDEEDETNCESTNLLKP
 join(34837, .35014, 35188, .35398, 35512, .36769)
 /note="ORF_ID:dd_00653"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAL99332.1"
 /db_xref="GI:19807763"
 /translation="MSEIDNTDIKKEMDNKNNYRDPDNKELLKWTKKKRRSNDSHI
 NPYRYNPNFKLLANDNDNKKIIGDITGTCIFPLIGELPKNYLCPTLIRINLY
 WISQDLNKLILINDNDNKKIIGDITGTCIFPLIGELPKNYLCPTLIRINLY
 LEYAQNITINSNLKILFKNEKSDILLKLLNYKEGNTFNSNDHDDHDDDD
 DEYFADFCINCPFPFKDLNENNNNNKPKSCTGVNEMVTDGGEFEFVKRIKES
 FOLKCKIRPYTTHIGKYNLPLINILIKOYILPKNOIQTTELVOGTSRWLWSYFL
 NKSTLETENNN
 IVGSSIKLDRDIEFLTFIDLTFLILFKPIDPKINNNNNNNNNNNNNNNNNNN
 NNNNNNNNNNSDGEIINSNLFILFKLENIKNEIKLK"
 37054, .38214
 /note="ORF_ID:dd_01574"
 /codon_start=1
 /product="PAP-1, 6/101"
 /protein_id="AAL99333.1"
 /db_xref="GI:19807764"
 /translation="MDRDYKSEKVKYGVNDKDRFSDNKKIRHNFDEFTMPNDRDR
 DKDIDRYDR
 NNNSSSSFDGDKYKRDMDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR
 GRSKRNTKYLEMVASNGNOKRLEDSDIPDPENREAREFLRNPATGGLHPLGK
 EVKVMQCKCFYGHRAICTQTNPIAEYELKREDPMIOYNDHDDDFGNNNNNIN
 NNNNNNNFIDKDRINKINTNDENNNDNNNNNNNNNNNNNNNNNNNNNNNN
 EKKRKRKKKIKASKSHRHSDDSDSGGSSDSDSDSLKDESYSSNESE"
 join(39790, 40108, 40252, .47744)
 /note="ORF_ID:dd_01572"
 /codon_start=1
 /product="Putative Acyltransferase"
 /protein_id="AAL99334.1"

CDS

CDS

CDS

Query Match 7.7% Score 44.8; DB 2; Length 105470;
 Best Local Similarity 51.5% Pred. No. 0.47; 97; Indels 0; Gaps 0;
 Matches 103; Conservative 0; Mismatches 0;
 Qy 376 CATAGAAGTAGAAAAATGAACACTGATTTAAACAAAGAGAGAGTAGCTACAAATGTGCC 435
 Db 4177 CACAAATATTAGATATTTTCGTACACACTCATCCACACACAAAAATATCTCTGATTA 4118
 Qy 436 TAGATATGACAAATATATCAAGAAACAGAGTACTGTCATAATAACTAACGACATGG 495
 Db 4117 TAATATTGTAATAATAATAACAAAAACGATATGTTAGAAAATAAAAAACATATTC 4058
 Qy 496 GACGAGTGAATAAACACTATATCTGTGCTGCTGTGAATTTGCTTATAGTATTTT 555
 Db 4057 ATCCAAATTTCCACCACTATGTTTGTGTTTTCCTACTACAAAAAATTTATATTCATTT 3998
 Qy 556 TGTCTTGATTTTCATTTTAC 575
 Db 3997 TTTTITTTTTTTTTTAAAC 3978

RESULT 11
 AX281294/c
 LOCUS AX281294 6641 bp DNA linear PAT 02-NOV-2001
 DEFINITION Sequence 36 from Patent W0017164.
 ACCESSION AX281294
 VERSION AX281294.1 GI:16608549
 KEYWORDS

SOURCE synthetic construct.
 ORGANISM synthetic construct
 REFERENCE 1
 AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
 TITLE Diagnosis of diseases associated with apoptosis
 JOURNAL Patent: WO 0177164-A 36 18-OCT-2001;
 Epigenomics AG (DE)
 FEATURES Location/Qualifiers
 source 1..6641
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"
 BASE COUNT 2495 a 22 c 984 g 3140 t
 ORIGIN
 Query Match 7.6% Score 44.4; DB 6; Length 6641;
 Best Local Similarity 48.8% Pred. No. 0.6;
 Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 Qy 333 AAACGATAAAATCGGTGGGACAAATAATAGTCTACCAACGCCATAAGAGTAGAAAAA 392
 Db 319 AAATATAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 260
 Qy 393 TATGAACACTGATTTAAACAAAGAGAGAGTAGCTACAAATTTGTCCTAGATATGACAATAT 452
 Db 259 TTCCCAATACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 200
 Qy 453 ATCAAGAAACAGGTACTATGTCATAATAACTAACGACATGGAGAGTAGGATAAAAAACA 512
 Db 199 TACTAAACACAATCTAATATAAATTAATAAATAAATAAATAAATAAATAAATAAATAA 140
 Qy 513 CTATATGCTGTCGTCGCTGTTGTAATTTGCTTAATAGTATTTTGTCTTGGATTTCATTT 572
 Db 139 ATATATATTTATTCGTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 80
 Qy 573 TACCAA 578
 Db 79 TACAAA 74
 RESULT 12
 AX345217/c
 LOCUS AX345217 6641 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 288 from Patent W00200928.
 ACCESSION AX345217
 VERSION AX345217.1 GI:18493103
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct
 REFERENCE 1
 AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
 TITLE Diagnosis of diseases associated with the immune system
 JOURNAL Patent: WO 0200928-A 288 03-JAN-2002;
 Epigenomics AG (DE)
 FEATURES Location/Qualifiers
 source 1..6641
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"
 BASE COUNT 2495 a 22 c 984 g 3140 t
 ORIGIN
 Query Match 7.6% Score 44.4; DB 6; Length 6641;
 Best Local Similarity 48.8% Pred. No. 0.6;
 Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 Qy 333 AAACGATAAAATCGGTGGGACAAATAATAGTCTACCAACGCCATAAGAGTAGAAAAA 392
 Db 319 AAATATAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 260
 Qy 393 TATGAACACTGATTTAAACAAAGAGAGAGTAGCTACAAATTTGTCCTAGATATGACAATAT 452

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5981
 Center clone name: 508_F1

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 167075 bases at least Q40
 Consensus quality: 169894 bases at least Q30
 Consensus quality: 170848 bases at least Q20
 Insert size: 174000; agarose-fp
 Quality coverage: 5.0 in Q20 bases; agarose-fp
 Quality coverage: 5.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1896: contig of 1896 bp in length
 * 1897 1996: gap of 100 bp
 * 1997 11367: contig of 9371 bp in length
 * 11368 11467: gap of 100 bp
 * 11468 20270: contig of 8803 bp in length
 * 20271 20370: gap of 100 bp
 * 20371 35910: contig of 15540 bp in length
 * 35911 36010: gap of 100 bp
 * 36011 57961: contig of 21951 bp in length
 * 57962 58061: gap of 100 bp
 * 58062 79942: contig of 21881 bp in length
 * 79943 80042: gap of 100 bp
 * 80043 118979: contig of 38937 bp in length
 * 118980 119079: gap of 100 bp
 * 119080 151341: contig of 32262 bp in length
 * 151342 151441: gap of 100 bp
 * 151442 172311: contig of 20870 bp in length.

FEATURES

source

1. .172311
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /map="14"
 /clone="RP11-508F1"
 /clone_lib="RPC1-11 Human Male BAC"
 misc_feature 1. .1896
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 1997. .11367
 /note="assembly_fragment"
 misc_feature 11468. .20270
 /note="assembly_fragment"
 misc_feature 20371. .35910
 /note="assembly_fragment"
 misc_feature 36011. .57961
 /note="assembly_fragment"
 misc_feature 58062. .79942
 /note="assembly_fragment"
 misc_feature 80043. .118979
 /note="assembly_fragment"
 misc_feature 119080. .151341
 /note="assembly_fragment"
 misc_feature 151442. .172311
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 BASE COUNT 54672 a 31706 c 31737 g 53396 t 800 others

ORIGIN

Query Match 7.6%; Score 44.4; DB 2; Length 172311;
 Best Local Similarity 49.6%; Prid No. 0.6;
 Matches 114; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
 QY 352 GACAATAAATAGTGTACCAAGCCCATAGAGAGTAAAGAAATATGAACCTGATTAAACAA 411
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 14536 GAAATAATACAGTCTACAGAGGCAAGAAAGCAAGCAAGCTAAACAAACAAAC 14595
 QY 412 AGAGAGTAGTACGAATTTGTCCTAGATATGCAATATATATCAAGAAGCAAGGTACTA 471
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 14596 AAACAATAAACCACCAAAACCCTGATGATGAAAGAACACATATAAGATACCGCAATA 14655
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 472 TGCATAAATAACTTAACGACGAATGGGACGAGTGATATAAAGACACTATATGTCTGTGGCTCGC 531
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 14656 TAATCCACATACCACACAGTCACACTAATATAATTTTAAAGATGCAAAATTTAA 14715
 QY 532 TTGTAATTTGCTTAATAGTATTTTCTTTGGATTTTCATTTTACCATAAT 581
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 14716 AAACTTTCAGATTAAAGTTATTACCTGTACATTCGTTTTTGAATAAT 14765

RESULT 15

CNS000008 218956 bp DNA linear PRI 26-APR-2001
 LOCUS Human chromosome 14 DNA sequence BAC R-330019 of library RPC1-11
 DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
 ACCESSION AL049831
 VERSION AL049831.2 GI:4972124
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 218956)
 AUTHORS Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
 Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
 Levy, M., Eckenberg, R., Bruls, T., DeBerardinis, V., Cruaud, C.,
 Gyapay, G., Saurin, W. and Weissenbach, J.
 TITLE Sequencing of the human chromosome 14
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 218956)
 Genoscope.
 DIRECT SUBMISSION
 TITLE Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
 JOURNAL Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr
 On Jun 2, 1999 this sequence version replaced gi:4808148.
 ----- Genome Center
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: <http://www.genoscope.cns.fr/>
 Contact: SeqRef@genoscope.cns.fr

 The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-850C20 (AC=AL121841)
 Downstream BAC (overlapping the SP6 end) : C-2216L14
 ----- Summary Statistics
 Assembly program: Phrap; version 2.0
 Quality coverage: 6.44x in Q20 bases; sum-of-contigs

 Overall quality chart :
 Range : bases
 0 :
 1 - 9 :
 10 - 19 :
 20 - 29 : 8
 30 - 39 : 13
 40 - 49 : 564
 50 - 59 : 2990
 60 - 69 : 3798

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 : Search time 136.703 Seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-12

Perfect score: 582

Sequence: 1 tacaataacgaataacta.....gatttcattttaccaatatt 582

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_101002:*
- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
 - 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
 - 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
 - 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
 - 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
 - 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
 - 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
 - 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
 - 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
 - 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
 - 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
 - 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
 - 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
 - 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
 - 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
 - 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
 - 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
 - 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
 - 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
 - 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
 - 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
 - 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
 - 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
 - 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	46.6	8.0	12025	ABL33298	Human immune syste
C 2	44.4	7.6	6641	ABL54336	Chemically treated
C 3	44.4	7.6	6641	ABL32315	Human immune syste
C 4	40.4	6.9	739	ABO65485	Arabidopsis thalia
C 5	40	6.9	7511	ABL33283	Human immune syste
C 6	39.8	6.8	18855	ABL32611	Human immune syste
C 7	39.6	6.8	542	ABQ28060	Oligonucleotide fo
C 8	39.6	6.8	542	ABQ28061	Oligonucleotide fo
C 9	39.6	6.8	5442	ABL33968	Human immune syste

C 10	39.4	6.8	6802	24	ABL33231	Human immune syste
C 11	39.2	6.7	6130	22	AAS46645	Tumour suppressor
C 12	39.2	6.7	6130	24	ABL32128	Human immune syste
C 13	39	6.7	5379	24	ABL70370	Chemically treated
C 14	39	6.7	5379	24	ABL33677	Human immune syste
C 15	39	6.7	5379	24	ABL34577	Human immune syste
C 16	38.8	6.7	840	21	ABN81104	Shrimp polynucleot
C 17	38.6	6.6	1182	22	ABA82969	Enterococcus faeca
C 18	38.6	6.6	4234	20	AA13250	Enterococcus faeca
C 19	38.4	6.6	1201	24	ABQ14374	Oligonucleotide fo
C 20	38.4	6.6	1201	24	ABQ14375	Oligonucleotide fo
C 21	38.4	6.6	2596	24	ABN59776	Novel human coding
C 22	38.4	6.6	7667	22	AAS46334	Tumour suppressor
C 23	38.4	6.6	18357	24	ABO67083	Human angioogenesis
C 24	38	6.5	788	22	AAH06825	Human CDNA clone (
C 25	38	6.5	2026	22	AAH17925	Human CDNA sequenc
C 26	38	6.5	6243	20	AAZ09598	Clostridium specie
C 27	38	6.5	11944	24	ABL34186	Human immune syste
C 28	37.8	6.5	12356	22	AAS46510	Tumour suppressor
C 29	37.8	6.5	14920	24	ABN80146	Human chemically m
C 30	37.6	6.5	11996	24	ABL34492	Human metastasis a
C 31	37.6	6.5	83391	24	ABO67094	Human angioogenesis
C 32	37.4	6.4	5153	24	ABL32955	Human immune syste
C 33	37.4	6.4	6775	24	ABO67159	Human angioogenesis
C 34	37.4	6.4	1664976	19	AAV21209	Methanococcus jann
C 35	37.2	6.4	898	12	AAQ11132	Skin corn formatio
C 36	37.2	6.4	5647	24	ABL70355	Chemically treated
C 37	37.2	6.4	5647	24	ABL33566	Human immune syste
C 38	37.2	6.4	5647	24	AA561320	Human gene regulat
C 39	37.2	6.4	6306	22	AAS45516	Chemically pretrea
C 40	37.2	6.4	6306	24	ABK28458	DNA transcription
C 41	37.2	6.4	6668	24	ABL33697	Human immune syste
C 42	37.2	6.4	19087	24	ABL32793	Human immune syste
C 43	37.2	6.4	38342	22	AAS46745	Tumour suppressor
C 44	37.2	6.4	38342	24	ABK31506	Signal transductio
C 45	37	6.4	5695	22	AAS46537	Tumour suppressor

ALIGNMENTS

RESULT 1

ABL33298/C

ID ABL33298 standard; DNA; 12025 BP.

AC ABL33298;

XX

XX 26-MAR-2002 (first entry)

DT Human immune system associated gene SEQ ID NO: 1271.

DE Human; immune system disease; cytosine methylation; antiasthmatic;

KW antihistaminic; antianemic; cytostatic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antihemetic; antiarthritic; antidiabetic; antipsoriatic;

KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;...

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

XX Homo sapiens.

OS WO200200528-A2.

PN 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

PA

XX

PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 1271; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 12025 BP; 3767 A; 84 C; 2123 G; 6051 T; 0 other;
 Query Match 8.0%; Score 46.6; DB 24; Length 12025;
 Best Local Similarity 58.2%; Pred. No. 0.0054;
 Matches 82; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 QY 380 AGAAGTAGAAAAATATGAACGTATTTAAACAAGAGAGAGTAGCTACAATTTGCCCTAGA 439
 DB 1407 ATAATATAAAAAATATAAATTTATTTAAATATAAAAAACCAATTTATTTTAAATAAT 1348
 QY 440 TATGACATATATATCAAGAAACAGGTACTATGCTAATATACTACAGCAATGGGACG 499
 DB 1347 TACCAAAATATAAAAAATATAAATTTATTTAAATATAAAAAATATAAAAAATATAAATA 1288
 QY 500 AGTGATAAAAAACACTATATGT 520
 DB 1287 AATAATACAAAACAAATATTT 1267
 RESULT 2
 ABL54336/c
 ID ABL54336 standard; DNA; 6641 BP.
 AC ABL54336;
 XX
 DT 29-JUL-2002 (first entry)
 DE Chemically treated apoptosis gene complementary to gene #18.
 XX
 KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
 KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
 KW amyotrophic lateral sclerosis; cancer; ds.
 XX
 OS Unidentified.
 XX
 PN WO200177164-A2.
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03969.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-017444/02.
 XX
 PT Chemically modified sequences of genes associated with apoptosis are
 PT useful to determine methylation patterns of genomic DNA samples for

PT diagnosis of associated diseases such as cancer
 XX
 PS Claim 1; Seq ID #36; 24pp; English.
 XX
 CC This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented in the printed specification but is based on
 CC information supplied by the European patent office.
 XX
 SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;
 Query Match 7.6%; Score 44.4; DB 24; Length 6641;
 Best Local Similarity 48.8%; Pred. No. 0.018;
 Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 QY 333 AAACGATATAAATGCGTGGGACAAATAAATAGTCTACCAACGCCATAGAAGTAGAAAAA 392
 DB 319 AACTATATAACAAAAAATAAATAAATAAACAACAAATACCAAAAAATTAACATTA 260
 QY 393 TATGAACGTATTTAAACAAGAGAGAGTAGCTACAATTTGTCCTAGATATGACAATATAT 452
 DB 259 TTCCCAATACCACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 200
 QY 453 ATCAACAACAGGTACTATGCTAATAAATACTACACCAATGGGACGAGTGAATAAACA 512
 DB 199 TACTAAACACAATCTAATATATAAATTTAATAAATAAATAAATAAATAAATAAATAA 140
 QY 513 CTATATCTGTGCTCGCTTGTAAATTTGCTTAATAGTTTGTCTTGGATTTCATTT 572
 DB 139 ATATATATTATTCGCTAAACACATAAATAAATAAATAAATAAATAAATAAATAAATA 80
 QY 573 TACCAA 578
 DB 79 TACAAA 74
 RESULT 3
 ABL32315/c
 ID ABL32315 standard; DNA; 6641 BP.
 AC ABL32315;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 288.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.

XX Olex A, Piepenbrock C., Berlin K;
 XX WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX Claim 1; SEQ ID NO 288; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;
 Query Match 7.68; Score 44.4; DB 24; Length 6641;
 Best Local Similarity 48.8%; Pred. No. 0.018;
 Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 Qy 333 AAACGATAAATGCGTGGGGACAATAAATAGTGTCTACCAAGCCCATAGAGTAGAAAAA 392
 Db 319 AAATATTAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 260
 Qy 393 TATGAATGATTTTAAACAAAGAGAGAGTAGCTACAATTTGTCCTAGATATGACAAATATAT 452
 Db 259 TTCCCAATACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 200
 Qy 453 ATCAAGAAACAGTACTATGTCATAAATACTACACCAATGGGACGAGTGATAAAAACA 512
 Db 199 TACTAAACACAACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 140
 Qy 513 CTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
 Db 139 ATATATATATTCCTCAAAACATATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 80
 Qy 573 TACCAA 578
 Db 79 TACAAA 74
 RESULT 4
 AB065485
 XX AB065485 standard; DNA; 739 BP.
 XX AC AB065485;
 XX DT 21-AUG-2002 (first entry)
 XX Arabidopsis thaliana polynucleotide SEQ ID NO 62.
 DE Arabidopsis thaliana; thale cress; plant; transgenic; GWO; disease;
 KW stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;
 KW insecticide; antibiotic; ds.
 XX OS Arabidopsis thaliana.
 XX PN US2002059663-A1.
 XX PD 16-MAY-2002.
 XX PF 26-JAN-2001; 2001US-0770149.
 XX PR 27-JAN-2000; 2000US-178506P.
 XX (GORLACH J.
 PA (ANY)/ AN Y.

PA (HAMI/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (RAIN/) RAINES T M.
 PA (YUY/) YU Y.
 PA (RAME/) RAMEAKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHAW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 PA (KRICK/) KRICKER M.
 PA (SLAT/) SLATER T.
 PA (DAVI/) DAVIS K R.
 PA (ALLE/) ALLEN K.
 PA (HOFF/) HOFFMAN N.
 PA (HURB/) HURBAN P.
 XX
 PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 PI Hurban P;
 XX WPI; 2002-479224/51.
 DR New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
 XX useful e.g. for preparing transgenic plants with increased resistance
 PT or altered metabolism
 PT
 XX Claim 1; SEQ ID NO 62; 40pp + Sequence Listing; English.
 PS The invention relates to nucleic acids (I) that hybridise under stringent
 CC conditions to any of 999 sequences (AB065424-AB0656422) or their
 CC fragments. (I) are used to express the corresponding polypeptides (II) or
 CC to produce genetically modified plant cells or transgenic plants, which
 CC may have improved resistance to disease or stress, or altered
 CC metabolic/biosynthetic pathways (for production of commercial,
 CC nutritional or medicinal products), or generally any trait of interest,
 CC or can be used to screen for biologically active agents (e.g. fungicides,
 CC insecticides and antibiotics)
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO at seqdata.uspto.gov/sequence.html?DocID=999909770149.
 XX
 SQ Sequence 739 BP; 289 A; 163 C; 134 G; 146 T; 7 other;
 Query Match 6.9%; Score 40.4; DB 24; Length 739;
 Best Local Similarity 53.9%; Pred. No. 0.11;
 Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 Qy 364 TGCTACCAAGCCCATAGAGTAGTAAATAATATCACTGATTAAACAAAGAGAGTAGC 423
 Db 286 TGCTATAAATCCAGCAAACTTGGGTAAAGATAAATTAATTAATGAAGAGTCCATAAC 345
 Qy 424 TACAATTTGCTCCCTAGATATGACAAATATATATCAAGAAACAGTACTATGCATAAATAAC 483
 Db 346 TGATGGTCTACCATATTGAAACTCTAAAACCATAGACCCAGCAGCAAGGCAATTA 405
 Qy 484 TACACCAATGGGACGAGTGATATAAACACTATA 517
 Db 406 CATCAGCATAGGAAAGAGCTATATATCAACAAA 439
 RESULT 5
 ABL33283/C
 ID ABL33283 standard; DNA; 7511 BP.
 XX
 XX ABL33283;
 XX AC
 XX DT 26-MAR-2002 (first entry)
 XX Human immune system associated gene SEQ ID NO: 1256.
 XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine; cytosine; cytosine;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS WO200200928-A2.
 PN 03-JAN-2002.
 PD 02-JUL-2001; 2001WO-EP07537.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX Claim 1; SEQ ID NO 1256; 32pp + Sequence Listing; German.
 PS The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX Sequence 7511 BP; 2046 A; 93 C; 1664 G; 3708 T; 0 other;

Query Match 6.9%; Score 40; DB 24; Length 7511;
 Best Local Similarity 51.7%; Pred. No. 0.37;
 Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
 OY 406 AACAAAGAGAGAGTACAAATGTCCTAGATATGACAAATATATATCAAGAAACAG 465
 Db 6546 AACCTAAACGACAAACGAACTCCGTCTCAAAAAAATAAAAAAATAAAAAA 6487
 OY 466 GTACTATGCATAATAACTACACCAATGGGACGAGTGATATAAACACTATATCTGTG 525
 Db 6486 AACCCCAATATAATTAACATTAACCAACCAAAAAAATAAAAAAATAAAAA 6427
 OY 526 GCTCGCTGTAATGCTATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 581
 Db 6426 TTCTACTTTTATTTCTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6371

RESULT 6
 ABL32611/c
 ID ABL32611 standard; DNA; 18855 BP.
 AC ABL32611;
 XX 26-MAR-2002 (first entry)
 DT Human immune system associated gene SEQ ID NO: 584.
 DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine; cytosine; cytosine;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS WO200200928-A2.
 PN 03-JAN-2002.
 PD 02-JUL-2001; 2001WO-EP07537.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX Claim 1; SEQ ID NO 584; 32pp + Sequence Listing; German.
 PS The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX Sequence 18855 BP; 5332 A; 178 C; 4053 G; 9292 T; 0 other;

Query Match 6.8%; Score 39.8; DB 24; Length 18855;
 Best Local Similarity 48.9%; Pred. No. 0.61;
 Matches 107; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
 OY 355 AATAAATAGTGTACCAACGCCATAGAGTACAAATAATGATGATTAACAAAGA 414
 Db 11201 ACTAATCCATTAACCTTTACAAACAAACCTTAAATAAATAAATCTCCCTCCACATATA 11142
 OY 415 GAGAGTAGCTACAAATGTCCTAGATATGACAAATATATATCAAGAAACAGGTACTATGC 474
 Db 11141 AACAAATCAATAATATATCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 11082
 OY 475 ATAAATAACTAACGCAATGGGACGAGTGATATAAACACTATATGCTGTGGCTCGCTTG 534
 Db 11081 TATACTAATTTTTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 11022
 OY 535 TAATTTGCTTAATAGTATTTTCTGTTGATTTTCATTTT 573
 Db 11021 TAAATTTCCAAATTT 10983

RESULT 7
 ABQ28060/c
 ID ABQ28060 standard; DNA; 542 BP.
 XX ABQ28060;
 AC ABQ28060;
 XX 12-JUL-2002 (first entry)
 DT Oligonucleotide for detecting cytosine methylation SEQ ID NO 14651.
 DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;
SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

XX Claim 12: 56pp + Sequence Listing: 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 542 BP; 45 A; 67 C; 201 G; 229 T; 0 other;

Query Match 6.8%; Score 39.6; DB 24; Length 542;
Best Local Similarity 46.2%; Pred. No. 0.17;
Matches 132; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 208 AACGGAAGAGTACGTACAACTATTAAGTGGGGAAGACGTCTCAGAACCCCAAAAT 267

DB 378 AACCGACCCCAACCGGAATAATCTAACGGGACACGCGACCCCGCACACCGCGCTCCG 319

QY 268 CCGGTGCACAAAGGTTAGAACGGTTTAAATTTCCGCCATTTTCGGCACCGCTGTAGTCCT 327

DB 318 CCCCACCCACATACCGCGACGAATAACCGCCCGCCCGCGGTAAACCAACGGGATT 259

QY 328 CCGCAAAACGATAAATGCGTGGGACATAAATAGTGTACCAACGCCCAAGAGTAG 387

DB 258 CCGCAAAACCCCAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 199

QY 388 AAAATATGACTGATTTAAACAAAGAGAGTAGTGTACAAATTTGCTCCTAGATGACAA 447

DB 198 AAAAAACCGGACCGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 139

QY 448 TATATATCAAGAACAGGTACTATGCTAATAAATTAACCTAACAGCAAT 493

DB 138 AAACCAACAAAAAAGAAAAAAGAAAAAATCCGCAACGACGACATAT 93

RESULT 8

ABQ28061

XX ID ABQ28061 standard; DNA; 542 BP.

XX AC ABQ28061;

XX DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 14652.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

XX Claim 12: 56pp + Sequence Listing: 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 542 BP; 229 A; 201 C; 67 G; 45 T; 0 other;

Query Match 6.8%; Score 39.6; DB 24; Length 542;

Best Local Similarity 46.2%; Pred. No. 0.17;

Matches 132; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 208 AACGGAAGAGTACGTACAACTATTAAGTGGGGAAGACGTCTCAGAACCCCAAAAT 267

DB 378 AACCGACCCCAACCGGAATAATCTAACGGGACACGCGACCCCGCACACCGCGCTCCG 224

QY 268 CCGGTGCACAAAGGTTAGAACGGTTTAAATTTCCGCCATTTTCGGCACCGCTGTAGTCCT 327

DB 318 CCCCACCCACATACCGCGACGAATAACCGCCCGCCCGCGGTAAACCAACGGGATT 284

Db 6273 ATCTAAAAAACTTAAACAAACAAAAATAACTATTCAATACAAAAAATAAAAAAAT 6214

Qy 394 ATGAACGATTTAAACAAAGAGAGAGTAGCTACAAATGTCCTAGATATGACATATATA 453

Db 6213 AATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6154

Qy 454 TCAAGAAACAGGTACTTACATATAAATAAATAACAGCAATGGGACGAGTGATAAAAA 510

Db 6153 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6097

RESULT 11

AAS46645/c

ID AAS46645 standard; DNA; 6130 BP.

AC AAS46645;

XX

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #367.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;

KW cytosine methylation; ds.

OS Homo sapiens.

XX WO200168912-A2.

PN

PD 20-SEP-2001.

XX

PF 15-MAR-2001; 2001WO-EP02955.

XX

PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX

PA (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2001-602752/68.

XX

PT Fragments of chemically modified genes associated with tumour suppressor

PT genes and oncogenes, useful in designing primers and probes for

PT analysing diseases associated with cytosine methylation state e.g.

PT cancer

XX

PS Claim 1; SEQ ID NO 367; 27pp; English.

XX

CC The invention relates to a nucleic acid comprising a sequence of 18

CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and

CC oncogenes having a sequence taken from 536 (actually 533 since

CC numbers 408, 458 and 500 are missing from the sequence listing) sequences

CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a

CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may

CC form part of a set of probes for detecting the cytosine methylation state

CC and/or single nucleotide polymorphisms and also to be used in an

CC array for analysing diseases associated with CpG dinucleotides e.g.

CC cancers and tumours. The probes can also be used in a method for

CC ascertaining genetic and/or epigenetic parameters for the diagnosis

CC and/or therapy of existing diseases or the predisposition to specific

CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the

CC differences serving as basis for diagnosis and/or prognosis events which

CC are disadvantageous to patients. The present sequence is one of the

CC 533 genomic sequences derived from tumour suppressor genes and

CC oncogenes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 6130 BP; 1489 A; 244 C; 1571 G; 2826 T; 0 other;

Query Match 6.7%; Score 39.2; DB 22; Length 6130;

Best Local Similarity 54.1%; Pred. No. 0.58;

Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 433 CCTAGATATGACAATATATATCAAGAAACAGGTACTATGTCATATAATACTAACACAA 492

Db 650 CTCAATATAACAAATAAAAAATAAAACACACAAAAAATAATACCGTACATCCA 591

Qy 493 TGGGACGAGTGATAAAAAACACTATATGCTGTGCTCGCTTGTGAATTTGCTTAATAGTTA 552

Db 590 AAGCATAAAAACATAAAAAACCATATTTCTATATTTATTACTATTATTATTTAATAAA 531

Qy 553 TTTTGTCTTGGATTTTCATTTTACCATA 580

Db 530 CTTTATTTTATTTTATTTTATTTTATTTTAAAAA 503

RESULT 12

ABL32128/c

ID ABL32128 standard; DNA; 6130 BP.

XX ABL32128;

AC

XX 26-MAR-2002 (first entry)

DT

DE Human immune system associated gene SEQ ID NO: 101.

XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antianemic; cytostatic; neutropenic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

XX

OS Homo sapiens.

XX WO200200528-A2.

XX

PD 03-JAN-2002.

XX

PF 02-JUL-2001; 2001WO-EP07537.

XX

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX

PA (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX

PT Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

XX

PS Claim 1; SEQ ID NO 101; 32pp + Sequence Listing; German.

XX

CC The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

xx The present invention provides a number of human immune system associated
cc genes which are modified by the methylation of cytosines. The sequences
cc can be used in the diagnosis and treatment of immune system disorders,
cc including eye diseases such as retinopathy, neovascular glaucoma and
cc macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
cc leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
cc rheumatoid arthritis, psoriasis and inflammatory/alcalterative bowel
cc diseases. The present sequence is a gene of the invention.

XX SQ Sequence 5379 BP; 1692 A; 52 C; 1038 G; 2597 T; 0 other;
 Query Match 6.7%; Score 39; DB 24; Length 5379;
 Best Local Similarity 48.4%; Pred. No. 0.63;
 Matches 108; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 360 ATAGTGTCTACCAACGCCATAGAAAGTAGAAAAATATGAACCTGATTTAAACAAGAGAGAG 419
 DB 2554 ATATAAATAAATTACTCTTCTAAATTTATTAAAAACAACAACACACATTAAACTAAATAAAT 2495

QY 420 TAGCTACAATTTGTCCTAGATATGACAATATATCAAAAGARACAGGTACTATGCATATA 479
 DB 2494 AAAATAAAAAATTAATAATCAATATCTCAATATCAATAAATAAATAAATAAATAAATA 2435

QY 480 TAACTACACAGCAATGGGACGAGTGATAAAAAACACATATATCTCTGCTCGCTTGTAAAT 539
 DB 2434 AAAATAAATCTATACCTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2375

QY 540 TGCTTAATAGTATTTTGTCTTGGATTTCATTTTACCAATATT 582
 DB 2374 TTCTAATCAATTATCTTTTAAAAATTTTCTCTCTCTTTCTT 2332

Search completed: December 24, 2002, 22:49:57
 Job time : 175.703 secs

RESULT 15
 ABL34577/c
 ID ABL34577 standard; DNA; 5379 BP.
 XX AC ABL34577;
 XX DT 26-MAR-2002 (first entry)
 XX DE Human metastasis associated gene SEQ ID NO: 130.
 XX KW Metastasis associated gene; cytostatic; gene therapy; cancer;
 XX KW cytosine methylation; gene; ds.
 XX OS Homo sapiens.
 XX PN WO200177376-A2.
 XX PD 18-OCT-2001.
 XX PF 06-APR-2001; 2001WO-EP03970.
 XX PR 06-APR-2000; 2000DE-1019058.
 XX PR 07-APR-2000; 2000DE-1019173.
 XX PR 30-JUN-2000; 2000DE-1032529.
 XX PR 01-SEP-2000; 2000DE-1043826.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX DR WPI; 2002-010922/01.
 XX PT New nucleic acid derived from chemically treated metastasis genes,
 XX PT useful for diagnosis of cancers by analysis of cytosine methylation,
 XX PT also for treatment
 XX PS Claim 1; SEQ ID NO 130; 23pp + Sequence Listing; English.
 XX CC The present invention provides a number of human metastasis associated
 XX CC genes which are modified by cytosine methylation. The sequences can be
 XX CC used in the diagnosis and treatment of cancer. The present sequence is
 XX CC one of the genes of the invention.
 XX SQ Sequence 5379 BP; 1692 A; 52 C; 1038 G; 2597 T; 0 other;

Query Match 6.7%; Score 39; DB 24; Length 5379;
 Best Local Similarity 48.4%; Pred. No. 0.63;
 Matches 108; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:16:24 ; Search time 3199.81 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-12

Perfect score: 582

Sequence: 1 tacaattaagaataaacta.....gatttcattttaccaatt 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estmu: *
4: em_estov: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_estl: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_othr: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	51.4	8.8	1030	CNS02LP3	AL203088 Tetraodon
C 2	48.2	8.3	889	CNS006W4	AL065999 Drosophil
C 3	46.8	8.0	777	CNS025WB	AL182612 Tetraodon
C 4	46.2	7.9	997	CNS0134P	AL102403 Drosophil
C 5	45.4	7.8	1201	CNS01660	AL106338 Drosophil
C 6	45.2	7.8	1125	9 AL547503	AL547503 AL547503

7	44.8	7.7	889	17	CNS006W4
8	44.6	7.7	639	17	CNS0170D
C 9	44.2	7.6	1101	17	CNS00L72
C 10	44	7.6	457	17	CNS009E1
C 11	44	7.6	891	17	AZ545935
12	43.6	7.5	877	17	AZ543123
13	43.6	7.5	894	17	AZ687792
C 14	43.4	7.5	836	17	AZ546944
C 15	43.4	7.5	880	17	BH155564
C 16	43.4	7.5	891	17	AZ539310
C 17	43.4	7.5	901	17	AZ548895
C 18	43.4	7.5	932	17	AZ689505
C 19	43.2	7.4	947	17	BH139082
C 20	43.2	7.4	1125	9	AL547503
21	43	7.4	875	17	CNS06MSO
22	43	7.4	928	17	CNS00DKY
C 23	42.8	7.4	832	17	CNS0110K
C 24	42.8	7.4	854	17	AZ688340
25	42.8	7.4	892	17	AZ540933
26	42.8	7.4	934	17	AZ542242
C 27	42.8	7.4	939	17	BH163549
C 28	42.6	7.3	898	17	AZ667815
C 29	42.4	7.3	417	17	CNS02AH7
30	42.4	7.3	1042	17	CNS00192
31	42.2	7.3	881	17	AZ673723
32	42.2	7.3	952	17	CNS006V8
C 33	42	7.2	804	17	CNS03GUX
C 34	42	7.2	847	17	AZ676214
35	42	7.2	917	17	AZ677473
C 36	42	7.2	951	17	AZ676519
37	42	7.2	1058	17	CNS011HR
38	42	7.2	1101	17	CNS00EVL
C 39	41.8	7.2	345	17	CNS04N9U
C 40	41.8	7.2	647	17	BH154473
41	41.8	7.2	799	17	AZ688611
C 42	41.8	7.2	822	17	AZ667519
C 43	41.8	7.2	828	17	AZ533304
44	41.8	7.2	828	17	CNS02PCX
45	41.8	7.2	830	17	AZ678784

ALIGNMENTS

RESULT 1	CNS02LP3	1030 bp	DNA	linear	GSS 14-MAY-2000
CNS02LP3/C	Tetraodon nigroviridis genome survey sequence T7 end of clone				
LOCUS	147L04 of library G from Tetraodon nigroviridis, genomic survey				
DEFINITION	Tetraodon nigroviridis genome survey sequence.				
ACCESSION	AL203088				
VERSION	AL203088.1 GI:7861433				
KEYWORDS	GSS: genome survey sequence.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Tetraodon nigroviridis				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes.				
AUTHORS	1 (bases 1 to 1030) Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Barnot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1030)				
AUTHORS	Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Barnot, A. and Weissenbach, J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis				
JOURNAL	Unpublished				

[illegible]

```

Qy 572 TTACCAATATT 582
   ||| |||
Db 594 TTATTTATTT 604

RESULT 8
CNS017QD 639 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN37M13 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108367
VERSION AL108367.1 GI:5628671
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 639)
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES             Location/Qualifiers
   source             1..639
   db_xref="taxon:7227"
   clone="BACN37M13"
   /clone_lib="DrosBAC"
   /plasmid="pBelobAC11"
   /note="end : SP6"
BASE COUNT 389 a 53 c 46 g 52 t 99 others
ORIGIN

Query Match 7.7% Score 44.6; DB 17; Length 639;
Best Local Similarity 38.08; Pred. No. 0.62; Indels 0; Gaps 0;
Matches 68; Conservative 36; Mismatches 75;

Qy 332 AAAACGATAAATGCGTGGGACATAAATAGTGTACCAACGCCATGAAGTAGAAAA 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 AWRRAAAWAAAWATATCAACWAAAWAAWARTAAWTAARAAWAAWATGGAGTAWAAA 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 392 ATATGAAGTATTTAAACAAGAGAGAGAGTAGTCAATGTCCTAGATATGACATATA 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 WATWTAATTTTAAAGAAAAAAWGTGGAARARARARAAWAAWAAWAAWAAATGAA 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 452 TATCAAGAACAGGTACTGTGTAATAATACTACACACATGGCAGCGATGATAAAAA 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 ARARWAGAGWNGTGGAAACACAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
CNS00LT2/c 1101 bp DNA linear GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL078714
VERSION AL078714.1 GI:5102004
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 457)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL

```

```

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Direct Submission
TITLE Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
   source             1..1101
   db_xref="taxon:7227"
   clone="BACR48P19"
   /clone_lib="RPCI-98"
   /note="end : TET3"
BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN

Query Match 7.6% Score 44.2; DB 17; Length 1101;
Best Local Similarity 40.38; Pred. No. 0.79; Indels 0; Gaps 0;
Matches 89; Conservative 29; Mismatches 103;

Qy 353 ACATAAATAGTGTACCAACGCCATGAAGTAGAAAAATATGAACGTATTAAACAAA 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 656 AAAAAAATWTAWAAWAAAAATTTWAAAAWAAAAWAAWAAWAAWAAWAAWAAWAA 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 413 GAGAGAGTAGCTACAATGTCCTCCATAGATATGACAAATATATATCAAGAACAGGTACTAT 472
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 WAAAAAAWATTAATAAAWTTTWTATMTMTMAWAAWAAWAAWAAWAAWAAWAAWAA 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 473 GCATAAATACTAACAGCAATGGCAGCGATGATAAACACATATATGCTCTGGCTCGCT 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 TWAAGAAWTAATAAAAHWTTTMMHHTTHAAWAAWAAWAAWAAWAAWAAWAAWAAW 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 533 TGTAAATTTGCTTAATAGTTATTTTGTCTTGGATTTCATTTT 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 TTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
CNS009EJ 457 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR19A04 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053475
VERSION AL053475.1 GI:4934951
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 457)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL

```

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammosier at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .457
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR19A04"
 /clone_lib="RPCI-98"
 /note="end : TET3"

BASE COUNT 181 a 51 c 58 g 62 t 105 others

ORIGIN

Query Match 7.68; Score 44; DB 17; Length 457;
 Best Local Similarity 36.68; Pred. No. 0.87; Indels 0; Gaps 0;
 Matches 68; Conservative 39; Mismatches 79; Indels 0; Gaps 0;
 QY 324 TCCTCCGCAACGATAAATGCGTGGGCAATTAATAGTGTACCAACGCCATAAGAA 383
 DB 220 TWMAGAAAMWTAAMAAWATWGTAAANWAGTGTGCGGCGRMAAAWAAAAA 279
 QY 384 GTAGAAAATGACTGATTAAACAAAGAGAGAGTACGATGTCCTCAGATAG 443
 DB 280 AAWAAAAWAWAGCTGGGTAAWAAAAWAAWAGWAAWMAWMAWMAWMAWMAW 339
 QY 444 ACAATATATACAAAGAACAGGTACTATGATTAATTAATCAACAGCAATGGGAGGTG 503
 DB 340 AAWAAAAWAAWAAWAAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAW 399
 QY 504 ATAAAA 509
 DB 400 AAWAAA 405

RESULT 11
 AZ545935/c
 LOCUS ENTFL49TR Entamoeba histolytica 891 bp DNA linear GSS 14-NOV-2000
 DEFINITION genomic, DNA sequence.
 ACCESSION AZ545935
 VERSION AZ545935.1 GI:11166985
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica.
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 891)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000).
 JOURNAL
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library

Seq primer: M13-Reverse
 Class: Shotgun
 High quality sequence start: 38
 High quality sequence stop: 706.
 Location/Qualifiers
 source 1. .891
 /organism="Entamoeba histolytica"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOSI; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. W. Vaudin and B.
 Barrell, Oxford University Press, 1999)."

FEATURES

source

277 a 95 c 80 g 439 t
 BASE COUNT
 ORIGIN

Query Match 7.68; Score 44; DB 17; Length 891;
 Best Local Similarity 51.08; Pred. No. 0.88;
 Matches 104; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 371 AACCCCATAGAGACGATGAATATGAATGATTTAAACAAAGAGAGTAGCTACAATT 430
 DB 545 AAAGAAATAGATAAATAGATAAGATGAATGAATGAATGAATGAATGAATGAAT 486
 QY 431 GTCCCTAGATAGACATATATATCAAGAACAGGTACTATGATCAATAAATCAACAGC 490
 DB 485 TCATTGAATTAATTCATTATGAATCAATGAATGAATGAATGAATGAATGAATGA 426
 QY 491 AATGGGACGAGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 550
 DB 425 AATAAAGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 366
 QY 551 TATTTTCTCTGGATTCATTTTA 574
 DB 365 ATTCTTATTTTATTTTCTATTTTA 342

RESULT 12
 AZ543123
 LOCUS ENTJ30TF Entamoeba histolytica 877 bp DNA linear GSS 14-NOV-2000
 DEFINITION genomic, DNA sequence.
 ACCESSION AZ543123
 VERSION AZ543123.1 GI:11161750
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica.
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 877)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000).
 JOURNAL
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward

Db 361 TAGTGCTACCAACGCCATAAGCAAGTAGAAAAATATGAAGTGAATTAACAAAGAGAGT 420
Qy 421 AGCTACAATTCCTCTAGATATGCAATATATCAAGAAACAGGTACTATGATCAAT 480
Db 421 AGCTACAATTCCTCTAGATATGCAATATATCAAGAAACAGGTACTATGATCAAT 480
Qy 481 AACTAACAGCAATGGGAGGAGTGATAAAACACATATATGTGTGGCTCGCTTGAATTT 540
Db 481 AACTAACAGCAATGGGAGGAGTGATAAAACACATATATGTGTGGCTCGCTTGAATTT 540
Qy 541 GCTTAATAGTATTTTGTCTGGATTCATTTTACCAATAT 582
Db 541 GCTTAATAGTATTTTGTCTGGATTCATTTTACCAATAT 582

RESULT 2

US-09-222-938A-15

; Sequence 15, Application US/09222938A

; Patent No. 6437108

; GENERAL INFORMATION:

; APPLICANT: Youngman, Philip

; APPLICANT: Fritz, Christian

; APPLICANT: Murphy, Christopher

; APPLICANT: Guzman, Luz-Maria

; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE

; FILE REFERENCE: 07334/060001

; CURRENT APPLICATION NUMBER: US/09/222.938A

; CURRENT FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 15

; LENGTH: 344

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-222-938A-15

Query Match 13.2%; Score 76.6; DB 4; Length 344;
Best Local Similarity 53.7%; Pred. No. 6.4e-14;
Matches 182; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

Qy 113 TGAATCCCGGATGTTAGTAGTAAAGGATGATACCCACATTTTCGACCAGCCAGGATATC 172
Db 6 TGTGACCTTGCTGGTGTGGAAGCGGTAAATCCATTTCTTCGACCATACCGTTGGAAC 65
Qy 173 GGCCTCTATAAACTTCCCTGTGACCGTTGACGTAAACGGAAGAGTACGTACAACAT 232
Db 66 ACTAACTGAAAAGTTTCTTGGGATCTTGGGACGAAAGGCTAATAAAAGTAGATGTC 125
Qy 233 ---AAGTGGGGAAGAACGTCCTCAGAAACGCCAAATCCGGTGACAAAGGAGTAGAAGC 289
Db 126 CCAAGAGGAGAGTAGAAGCTGAAACCCGACATACCGGTATGGAGGATAGAAAC 185
Qy 290 GGTTTAAATTTCCGCCATTTCCGCCACCGCTGTAGTCTCCGCAAAACGATAAAATGCGTG 349
Db 186 GTCTTAATTTCCACCAATTCGACAGCGTGTGTCACGACCTCACTAAAAGCCTAAACGCG 245
Qy 350 GGCACAAATAATAGTCTACCAACCCGATAGAGAGTACGAAATATGACTGATTTAAAC 409
Db 246 GATAGAGAGAGATGGAAGCGTAAATAGAGAAACCTTGAGAGTATAGAACCGTCATACT 305
Qy 410 AAAGAGAGATGAGTACAAATTTCTCCCTAGATATGACAAT 448
Db 306 AAAGTGACAGATCACAGTGTCTGCTAGCTAGCGCCGACAT 344

RESULT 3

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match. 8.2%; Score 48; DB 1; Length 7218;

Best Local Similarity 2.7%; Pred. No. 8.4e-05;

Matches 9; Conservative 194; Mismatches 129; Indels 0; Gaps 0;

Qy 203 GACGTACGGAAGAGAGTACGTACAACATATAGTGGGGAAGAGAGTCTCTCAGAAACGCC 262
Db 1382 RRR 1323
Qy 263 AAAATCCGGTGCACAAAGGAGTAGAAGCGTTTAAATTTCCGCCATTTCCGCCACCGCTGA 322
Db 1322 RRR 1263
Qy 323 GTCCTCCGCAAAACGATAAATGCGTGGGACAAATAAGTGTACCAACGCCATAAGA 382
Db 1262 RRR 1203
Qy 383 AGTAGAAAATATGAAGTATTTAAACAAGAGAGAGTAGTACAATTTGCCCTAGATAT 442
Db 1202 RRR 1143
Qy 443 GACAATATATCAAGAAACAGGACTATGCTATCAATAAATACTAAACAGCAATGGGACGAGT 502
Db 1142 RRR 1083
Qy 503 GATAAAACACTATATGTCTGTGCTGCTGCTTG 534
Db 1082 RRR 1051

RESULT 4

US-09-056-075-1/c

; Sequence 1, Application US/09056075
; Patent No. 5955368

GENERAL INFORMATION:

; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/056,075
; FILING DATE:

CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:

NAME/KEY: misc_feature

LOCATION: 3770..4013

; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from
; OTHER INFORMATION: plasmid RP4"

US-09-056-075-1

Query Match 6.5%; Score 38; DB 2; Length 6243;

Best Local Similarity 49.5%; Pred. No. 0.081;

Matches 98; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 327 TCCGCAAAACGATAAAATGCGTGGGACAAATAAATAGTGCTACCAACGCCCATAGAAGTA 386

Db 1715 TCTAAAGGATATACAGAGAGAGAAATATTAGAGGTAAATAATATATCAAGAGCAA 1656

Qy 387 GAAAAATATGAACGTGATTTAAACAAGAGAGAGTAGCTACAAATTTGCCCTAGATATGACA 446

Db 1655 TTAAGAGCTAAAGGAAAGCAACAAACAAAGAGGAATTAATGTATTAAAGAAAAAATA 1596

Qy 447 ATATATATCAAGAAACAGGTACTACTATGATTAATACTAACGCAATGGGACGAGTGATA 506

Db 1595 AAAGCCCTTAAGAAAGAAAGGCTTTAAATAATAAGAAATTAATCTAATGTTAGAAGTACCA 1536

Qy 507 AAACACTATATGTCGT 524

Db 1535 ATAAAAACATTGAACGT 1518

RESULT 5

US-08-353-341-4

; Sequence 4, Application US/08353341

; Patent No. 565417

GENERAL INFORMATION:

; APPLICANT: SUGINAKA, Hidekazu

; APPLICANT: SUGAI, Motoyuki
; APPLICANT: HON, Yonman
; APPLICANT: OGAI, Hideo
; TITLE OF INVENTION: POLYPEPTIDE, PROCESS FOR PREPARING THE
; TITLE OF INVENTION: SAME, AND PHARMACEUTICAL COMPOSITIONS AND COSMETICS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas P. SARRO
; STREET: 727 Twenty-Third Street
; CITY: South Arlington
; STATE: Virginia
; COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/353,341
; FILING DATE: 05-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 689252
; FILING DATE: 12-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: SARRO, Thomas P.
; REGISTRATION NUMBER: 19396
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-920-7200
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-353-341-4

Query Match 6.4%; Score 37.2; DB 1; Length 899;

Best Local Similarity 56.6%; Pred. No. 0.066;

Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 372 ACGCATAAGAGTAGAAAAATATGAAGTATTTAAACAAGAGAGAGTAGCTACAATTG 431

Db 754 ACGCTGTGGAGTGTAGATTGTCAAATGATAAAAGAAATCATATACACGCTATTG 813

Qy 432 TCCCTAGATATGACAATATATATCAAGAAACAGGTACTATGCTAATAACTAACACGCA 491

Db 814 TTTTAAAAAATACAAATATATAACAAGATATAACAAGGATTATAATAATTAATAATTTTA 873

Qy 492 AT 493

Db 874 AT 875

RESULT 6

US-09-134-001C-1065/c

; Sequence 1055, Application US/09134001C

; Patent No. 6380370

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674

ORGANISM: Plasmodium yoelii
US-08-929-329-1

```

RESULT 9
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA

```


; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1293 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE:
 ; ORIGINAL SOURCE:
 ; FEATURE:
 ; US-08-682-080-1

Query Match 5.6%; Score 32.6; DB 3; Length 1293;
 Best Local Similarity 52.6%; Pred. No. 1.8;
 Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 376 CATAAGAGTAGAAAATATGAACCTGATTTAAACAAGAGAGAGTAGCTACAAATGTCCTCC 435
 DB 963 CTTAGGACGTGAATATGCTGAGAAAACCTGAAAAGGTATAATATTAGAAATGTCAC 904
 QY 436 TAGATATGACAAATATATCAAGAAACAGTACTATGCATATAATACTAACACCATGG 495
 DB 903 TGTAGGACATGGAATATAGCAAGAAAACCTGACACTCATGGAATAATGAGAAACATCCCTT 844
 QY 496 GACGAGTGATAAAAA 510
 DB 843 GACAACTGGAAGAAA 829

RESULT 12
 US-08-487-826B-13
 ; Sequence 13, Application US/08487826B
 ; Patent No. 5993827

; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellens, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelsen, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CP1
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19124 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single

; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-487-826B-13

Query Match 5.6%; Score 32.6; DB 2; Length 19124;
 Best Local Similarity 48.2%; Pred. No. 5.3;
 Matches 92; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 392 ATATGAACCTGATTTAAACAAGAGAGAGTAGCTACAAATGTCCTCCCTAGATATGACAATATA 451
 DB 6285 ATATTTGAATATATATATTATAATGCTATTTTACTTATTTTATTTATATATACATA 6344
 QY 452 TATCAAGAAACAGGTACTATGCAATAAATAACGCAATGCGACGAGTGATAAAAAAC 511
 DB 6345 AAATTTTGAATTCATAAATGCAATGCAATAAATAAACAACAACAACAATGATA 6404
 QY 512 ACTATATGCTGTGGCTCGCTTCTAATTTGCTTAATAGTATTTTGTCTTGGATTTCATT 571
 DB 6405 AAAACATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6464
 QY 572 TTACCAATATT 582
 DB 6465 TTATCATTTT 6475

RESULT 13
 US-08-785-310A-3/C
 ; Sequence 3, Application US/08785310A
 ; Patent No. 5840532
 ; GENERAL INFORMATION:
 ; APPLICANT: McKnight, Steven L.
 ; APPLICANT: Russell, David W.
 ; TITLE OF INVENTION: Neuronal PAS Domain Protein
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 ; STREET: 268 BUSH STREET, SUITE 3200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/785,310A
 ; FILING DATE: 21-JAN-1997
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OSMAN, RICHARD A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: UTSD:1226
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 343-4341
 ; TELEFAX: (415) 343-4342
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4010 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; US-08-785-310A-3

Query Match 5.5%; Score 32.2; DB 2; Length 4010;
 Best Local Similarity 57.4%; Pred. No. 3.8;
 Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 393 TATCAACTGATTTAAACAAGAGAGAGTAGCTACAAATGTCCTCCCTAGATATGACAATATAT 452

Db 3313 TATTATTTTTTAAAAAACAAGAGTGCCTAAGTTATTTCCTCTGTAGTGTCGAA 3254
QY 453 ATCAAGAACAGGTACTATGCTATATAATAACTAACACCAAT 493
Db 3253 ATTAAACAACAACACCATTTTAAAAATAATGAATAATAAT 3213

RESULT 14
US-08-961-527-149
; Sequence 149, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 149:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-149

Query Match 5.5%; Score 32.2; DB 4; Length 12566;
Best Local Similarity 51.0%; Pred. No. 5.9;
Matches 76; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 343 ATCGGTGGGACAAATAATAGTCTACCAACGCCATAGAAGTAGAATAATATCAACTGA 402
Db 2281 AAGTGGGGCTTTAGAAAAAGTTAACTAGGTATTGGATTAGAACAAAAATCTATA 2340
QY 403 TTTAAACAAGACAGAGTAGCTACAAATTTGCTCCTAGATATGACAATATATATCAAGAAA 462
Db 2341 CTTTATCTGGGGAGAGGCGCCCAACAGAGTTGCGCTTGTAGACTATTTTGAATAATCCAC 2400

QY 463 CAGGTACTATGCATAATAACTAACAGCA 491
Db 2401 CCTTGATTTTGGCAGATGAACCAACAGCA 2429

RESULT 15
US-09-325-932A-16/c
; Sequence 16, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry

; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant d
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Pinus radiata
; US-09-325-932A-16

Query Match 5.4%; Score 31.4; DB 4; Length 530;
Best Local Similarity 48.1%; Pred. No. 3;
Matches 89; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 367 TACCAACGCCATAGAAGTAGAAAAATATGAACCTGATTTTAAACAAGAGAGAGTAGCTAC 426
Db 484 TACAAACATATTAAATAAGTAGTGATTTCACCTTAATATGCTGAAGGACTCGTTGATAGCTG 425
QY 427 AATTGTCCTAGATATGACAATATATATCAAGAAACAGGTACTATGCATAATAATACTAA 486
Db 424 ACAGATCCCCCAATATGAGAAAAATGTGGATGTAATAATGTACATATAAGCTATAACCAA 365
QY 487 CACCAATGGGACGAGTCATAAAACACTATATGCTGTGCTCGCTTGTAAATTTGCTTAA 546
Db 364 CACTAATAATACGATTGGCACATTAACACTGCTTCTGTATATATTGACAAAAATGTTCAA 305

QY 547 TAGTT 551
Db 304 AATT 300

Search completed: December 25, 2002, 14:27:11
Job time : 119.875 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 ; Search time 171.89 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-12
Perfect score: 582
Sequence: 1 tacaataacgaataacta.....gatttcatttaccatatt 582

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues
Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	582	100.0	582	12	US-10-068-080-12
2	100	17.2	642	12	US-10-068-080-11
3	40.4	6.9	739	10	US-09-770-149-62
4	38.6	6.6	4234	10	US-09-070-927A-313
5	37.4	6.4	392	10	US-09-960-352-11703
6	36.8	6.3	1985	9	US-09-905-291A-212
7	36.8	6.3	1985	9	US-09-902-853-212
8	36.8	6.3	1985	10	US-09-909-320-212
9	36.8	6.3	1985	10	US-09-909-088B-212
10	36.8	6.3	1985	12	US-10-052-586-27
11	36	6.2	832	9	US-10-001-857-97
12	36	6.2	27377	10	US-09-816-248-18
13	35.8	6.2	1170	9	US-09-938-842A-5002
14	35.6	6.1	376	10	US-09-960-352-5087
15	35.6	6.1	428	10	US-09-960-352-573
16	35.6	6.1	4167	10	US-09-764-878-282
17	35.6	6.1	4167	10	US-09-764-860-1145
18	35.6	6.1	4167	10	US-09-764-846-345
19	35.4	6.1	357	10	US-09-960-352-1692

c 20	35.4	6.1	446	10	US-09-960-352-3400	Sequence 3400, Ap
c 21	35.4	6.1	516	10	US-09-960-352-5785	Sequence 5785, Ap
c 22	35.4	6.1	4237	10	US-09-745-763-20	Sequence 20, Appl
c 23	35.4	6.1	21636	10	US-09-416-384A-3	Sequence 3, Appl1
c 24	35	6.0	414	10	US-09-960-352-6528	Sequence 6528, Ap
c 25	35	6.0	442	10	US-09-960-352-12911	Sequence 12911, A
c 26	35	6.0	480	10	US-09-960-352-5301	Sequence 5301, Ap
c 27	34.8	6.0	426	10	US-09-960-352-8406	Sequence 8406, Ap
c 28	34.8	6.0	463	10	US-09-960-352-7186	Sequence 7186, Ap
c 29	34.8	6.0	21860	10	US-09-764-877-3803	Sequence 3803, Ap
c 30	34.8	6.0	65359	10	US-09-804-472-3	Sequence 3, Appl1
c 31	34.6	5.9	385	10	US-09-960-352-1739	Sequence 1739, Ap
c 32	34.6	5.9	431	10	US-09-960-352-5558	Sequence 5558, Ap
c 33	34.4	5.9	377	10	US-09-960-352-5583	Sequence 5583, Ap
c 34	34.4	5.9	1198	10	US-09-789-561-55	Sequence 55, Appl
c 35	34.2	5.9	484	10	US-09-817-607-10	Sequence 10, Appl
c 36	34.2	5.9	1959	10	US-09-070-927A-282	Sequence 282, App
c 37	34	5.8	352	10	US-09-960-352-10129	Sequence 10129, A
c 38	34	5.8	1200	10	US-09-887-576-758	Sequence 758, App
c 39	34	5.8	2000	9	US-09-938-842A-3681	Sequence 3681, Ap
c 40	33.8	5.8	415	10	US-09-960-352-2223	Sequence 2223, Ap
c 41	33.8	5.8	448	10	US-09-960-352-11904	Sequence 11904, A
c 42	33.8	5.8	451	10	US-09-960-352-10262	Sequence 10262, A
c 43	33.6	5.8	948	9	US-09-938-842A-2359	Sequence 2359, Ap
c 44	33.6	5.8	259	10	US-09-960-352-7233	Sequence 7233, Ap
c 45	33.6	5.8	439	10	US-09-960-352-1009	Sequence 1009, Ap

ALIGNMENTS

RESULT 1
US-10-068-080-12
; Sequence 12, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-12

Query Match	100.0%	Score 582;	DB 12;	Length 582;
Best Local Similarity	100.0%;	Pred. No. 5.9e-157;		
Matches 582;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TACATTTACGAAATACTAATAAAACCGGATCAATATATCGTAAAGCTAGACCGAAT	60	
Db	1	TACATTTACGAAATACTAATAAAACCGGATCAATATATCGTAAAGCTAGACCGAAT	60	
QY	61	TAAACCCGTTGCGAAGCGTTTCCTTAACTATAAGCCCTCGTGGCTTGCCTGTTGAATCCG	120	
Db	61	TAAACCCGTTGCGAAGCGTTTCCTTAACTATAAGCCCTCGTGGCTTGCCTGTTGAATCCG	120	
QY	121	CGATGGTTAGCTAAGGATGTAACCCACATTTTCGACCAAGCAGTAGTATCGGCTCTTA	180	
Db	121	CGATGGTTAGCTAAGGATGTAACCCACATTTTCGACCAAGCAGTAGTATCGGCTCTTA	180	
QY	181	TAAACTTTCCCTGTGACCCCTTGACGTAACGGAAAGAGTACGTACAACTATAGTGGCG	240	
Db	181	TAAACTTTCCCTGTGACCCCTTGACGTAACGGAAAGAGTACGTACAACTATAGTGGCG	240	
QY	241	GAAGAACGTCCTCAGAAACCCCAAAATCCGGTCACAAAGGGTAGAAGCGGTTTAAATTT	300	

```

Db 241 GAAGAAGCTCCTCAGAAACGCCAAAATCCGGTGCACAAAGGGTAGAAGCGGTTTAAATTT 300
QY 301 CCGCCATTTCCGACCCCTCTAGTCTCCGCAAAAGCATAAATCGCTGGGGACAATAA 360
Db 301 CCGCCATTTCCGACCCCTCTAGTCTCCGCAAAAGCATAAATCGCTGGGGACAATAA 360
QY 361 TAGTGTCTACCAAGCCCAATAAGAAGTAGAAAAATATGAACGTGATTTAAACAAGAGAGAT 420
Db 361 TAGTGTCTACCAAGCCCAATAAGAAGTAGAAAAATATGAACGTGATTTAAACAAGAGAGAT 420
QY 421 AGCTACAATTTGCTCCCTAGATATGACAAATATATATCAAAAGAAACAGGTACTATGCAATAAT 480
Db 421 AGCTACAATTTGCTCCCTAGATATGACAAATATATATCAAAAGAAACAGGTACTATGCAATAAT 480
QY 481 AACTAACAGCAATGGGACGAGTGATAAAAACACTATATGCTGTGCTCCCTGCTTAATTT 540
Db 481 AACTAACAGCAATGGGACGAGTGATAAAAACACTATATGCTGTGCTCCCTGCTTAATTT 540
QY 541 GCTTAATAGTATTGCTCTTTGGATTTCATTTTACCATAATT 582
Db 541 GCTTAATAGTATTGCTCTTTGGATTTCATTTTACCATAATT 582

RESULT 2
US-10-068-080-11
; Sequence 11, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-11

Query Match 17.2%; Score 100; DB 12; Length 642;
Best Local Similarity 52.9%; Pred. No. 3.1e-19;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 1 TACAATTAACGAAATACTAATAAAACCGGATGAACATATCCGTCGTAAGGTAGACCGAAT 60
Db 1 TACTAATGTTATCAAAATATATAGGATCGGATAGACGACCCCAAGCTAAGGTAGACCAAG 60
QY 61 TAACACCGGTTGCAACGGTTTCCTTAATATAAGCCCTCGTGCTCGCTGCGGTTGAATCCG 120
Db 61 ACCTAAGCTGTTCAATAAGAAGTTTAGTTAGATCGCTCGCTGCTACCAAGACCAATTTGACCT 120
QY 121 CGATGGGTAGTAAAGCATGTACCCACATTTTCGACCAAGCCAGCAGATATCGCCCTCTA 180
Db 121 TCGTGGTTGTGGAAGGCGTAAATTCATCTCTTCGACCATACCGTTGGAACACATAACTG 180
QY 181 TAAAACTTTCCTGTGACCGTTGACGTAAAGGAAAGAGTACGTACA- --ACTATAAGTG 237
Db 181 AAAAGTTTCTTGGGATCGTTGGCAGCAAGGCTAATAAAAAGTAGATGTTCCCAAGA 240
QY 238 GCGAAGAACGCTCTCAGAAACGCCAAATCCGGTCGACAAAGGTTAGACGGTTTAA 297
Db 241 GGAGAGTAGAAACCTGAAAACCGCAATAGCCGGTATGGAAGGGATAGAAACGCTCTTAA 300
QY 298 TTTCCGCAATTTCCGACCGCTAGTCTCCCGCAAAACGATATAAATCGCTGGGCAAT 357
Db 301 TTTCCACCATTTCCGACAGCGTTGCTCAGCAGCTTCACCTATAAAGCCTAAACCGGATAGAAG 360
```

```

QY 358 AATAAGTGTCTACCAACGCCCATAGAAGTAGAAAAATATGAACGTGATTTAAACAAGAGAG 417
Db 361 ACAGAGATGGAACGCTAATAAGAAGAAACCTCGAGAGATAGAACCGTCACTACTAAAGTGAC 420
QY 418 AGTAGCTACANTTGTCCCTAGATATGACAATA 449
Db 421 AGATCACAGTGTGCTAGCTAAACGCCGACAATA 452

RESULT 3
US-09-770-149-62
; Sequence 62, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: Thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows, Version 4.0
; SEQ ID NO 62
; LENGTH: 739
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(739)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-149-62

Query Match 6.9%; Score 40.4; DB 10; Length 739;
Best Local Similarity 53.9%; Pred. No. 0.035;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 364 TGCTACCAAGCCCATAGAAGTAGAAAAATATGAACGTGATTTAAACAAGAGAGAGTAGC 423
Db 286 TGCATATAATTCACGCAAACTTGGTAAAGTAAATTTGATTAATTTGAAGAGTCCATAAC 345
QY 424 TACAATTTGCTCCCTAGATATGACAATATATATCAAAAGAAACAGGTACTATGCAATAAATAC 483
Db 346 TGATGGTCTACCATATTTGAAACTCTAAAAACCATAGAACCCAGCGCAAAAGCAATTAA 405
QY 484 TACACCAATGGGACGAGTGATATAAACACTATA 517
Db 406 CATCAGCATAGGAAAAAGAGCTATAATACCCAA 439

RESULT 4
US-09-070-927A-313
; Sequence 313, Application US/09070927A
; Patent No. US20020120116A1
```



```
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 212
; LENGTH: 1985
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-905-291A-212

Query Match      6.3%; Score 36.8; DB 9; Length 1985;
Best Local Similarity 51.9%; Pred. No. 0.58;
Matches 83; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 351 GGACATAAATAGTGTACCAAGCCATAGAGAGTAGAGAAATATGAACTGTTTAAACA 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1826 GGACATAAATAGTGTACCTTCGCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 411 AAGAGAGAGTAGTACAAATGTCCCTAGATATGACAAATATATATCAAGAAACAGGTACT 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1886 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1945
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 471 ATGCATAAATACTACAGCAATGGCAGCGATGATAAAAA 510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1946 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-09-902-853-212
; Sequence 212, Application US/09902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
```



```
US-09-909-320-212
; Sequence 212, Application US/09909320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; PRIOR FILING DATE: 2002-01-04
; CURRENT APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 212
; LENGTH: 1985
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-909-320-212
Query Match 6.3%; Score 36.8; DB 10; Length 1985;
Best Local Similarity 51.9%; Pred. No. 0.58; Mismatches, 77; Indels 0; Gaps 0;
Matches 83; Conservative 0;

QY 351 GGACAATAAATAGTGTACCAACGCCATAGAGAGTAAAGAAATATGAACTGATTTAAACA 410
Db 1826 GGACAATAAATAGTGTATGACTGCTTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1885
QY 411 AAGAGAGAGTAGCTACAAATTTGTCCTAGATATGACATATATATCAAGAAAGACAGGTACT 470
Db 1886 AAAAAAATAAATAGTGTATGACTGCTTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1945
QY 471 ATGCATAAATACTAAACAGCAATGGGACGAGTGATATAAAAA 510
Db 1946 AAAAAAATAAATAGTGTATGACTGCTTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1985

RESULT 9
US-09-909-089B-212
; Sequence 212, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
```

```

; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 212
; LENGTH: 1985
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-909-0889-212

Query Match      6.3%; Score 36.8; DB 10; Length 1985;
Best Local Similarity 51.9%; Pred. No. 0.58;
Matches 83; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

OY 351 GGACATAAATAGTGTACCAAGCCCATAGAGAGTACAGAAATATGAACTGATTTAAACA 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1826 GGACATAAATAGTGTACCTGCTGCTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1885

OY 411 AAGCAGAGTAGTACATGTTCCCTAGATATGACATATATATCAAGAAACAGGTACT 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1886 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1945

OY 471 ATGCATAAATACTACAGCAATGGACGAGTGATAAAA 510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1946 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1985

RESULT 10
US-10-052-586-27
; Sequence 27, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: F3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
```

```
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088722
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
```

```
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908

Query Match          6.3%; Score 36.8; DB 12; Length 1985;
Best Local Similarity 51.9%; Pred. No. 0.58; Mismatches 77; Indels 0; Gaps 0;
Matches 83; Conservative 0;

Qy 351 GGACATAAATAGTGCTACCAACGCCATAGAGAGTAAAGAAATATGCACTGATTTAAACA 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1826 GGACATAAATGGTGTATGCTGCTTCGCCCAAAAAAATAAAAAAATAAAAAA 1885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 411 AAGAGAGAGTAGCTACAATTTGTCCTTAGATATGACAATATATATCAAGAACAGGTACT 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1886 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1945
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 471 ATGCATAAATACTACACCAATGGGACGAGTGATAAAA 510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1946 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-10-001-857-97
; Sequence 97, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-97

Query Match          6.2%; Score 36; DB 9; Length 832;
Best Local Similarity 49.5%; Pred. No. 0.67; Mismatches 95; Indels 0; Gaps 0;
Matches 93; Conservative 0;

Qy 330 GCAAAACGATAAAATCGTGGGACAAATAAATAGTCTACCAACGCCATAAGAGTAGAA 389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 GGACACAGACAGAGAGTAAGGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 390 AATATGAATGATTTTAAACAAAGAGAGAGTAGCTACAATTTGCCCTAGATATGACAATA 449
```

D6	348	AAAAATATAAATAATACAAATAAACCAATTAATAAAAAAAAAAAAAAAAAAAAAATAAAAA
QY	392	ATATGAAGCTGATTAAAACAAGAGAGAGTAGCTACAAATTCCTCCCTAGCATATGCACAATATA
D6	288	TAAATAAATAATAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
QY	452	TATCAGAAGAACAGGTACTATGCATAAATAACTAACAGCAANTGGGACGAGTGATATAAAC
D6	228	AATTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
QY	512	ACTATA 517
D6	168	AAAAATA 163

RESULT 15
US-09-960-352-573/c
; Sequence 573, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LA

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
 ; FILE REFERENCE: 16511.006/37-21(10298)C
 ; CURRENT APPLICATION NUMBER: US/09/960,352
 ; CURRENT FILING DATE: 2001-09-24
 ; NUMBER OF SEQ ID NOS: 15112
 ; SEQ ID NO 573
 ; LENGTH: 428
 ; TYPE: DNA
 ; ORGANISM: Bos taurus
 ; OTHER INFORMATION: Clone ID: 03-LIB3058-031-Q1-K1-A11
 US-09-960-352-573

Query Match 6.1%; Score 35.6; DB 10; Length 428;
 Best Local Similarity 49.5%; Pred. No. 0.64;
 Matches 92; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy	332	AAACGATAAATCGCTGGGACAAATTAATAGTCTACCAACGCCATAAGAAGTAGAAAA	391
Db	376	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	317
Qy	392	ATATGAACCTGATTTAAACAAAGAGAGAGTAGCTACAATTTGCTCCTAGATATGACAATATA	451
Db	316	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	257
Qy	452	TATCAAGAAACAGGTACTATGCATAAATAACTAACAGCAATGGGACGAGTGATAAAAC	511
Db	256	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	197
Qy	512	ACTATA 517	
Db	196	AAGAAA 191	

Search completed: December 26, 2002, 00:46:35
 Job time : 201.89 secs